DOCUMENT RESUME

ED 029 958 VT 007 886

By-Morton, J.E.

Analytical Potential of the Current Population Survey for Manpower and Employment Research. Studies in Employment and Unemployment.

Upjohn (W.E.) Inst. for Employment Research, Kalamazoo, Mich.

Pub Date Dec 65

Note-73p.

Available from-W.E. Upjohn Institute for Employment Research, 1101 Seventeenth Street, N.W., Washington, D.C. 20036 (single copies without charge)

EDRS Price MF-$0.50 HC-$3.75

Descriptors-*Data Analysis, Data Processing, *Employment, Information Retrieval, Information Sources, Information Storage, Information Systems, *Information Utilization, Job Market, Labor Supply, *Manpower Needs, Records (Forms), *Research Tools

Identifiers-CPS, *Current Population Survey

The primary scope of this publication is to contribute to the promotion of research in the general area of manpower and employment by surveying the supply of statistical data with a view to stimulating and facilitating their analytical use. Since the mass of relevant manpower data in this country is enormous, it was decided to approach the problem in steps and to select the Current Population Survey as the first major data source for examination. To achieve this objective, the present study undertakes to stimulate analytical efforts by considering not only the content of the information, but also the format in which it might be usefully available for analysis: hence the emphasis on preparing the prospective data-consumer for some of the consequences of the enormous size and the increasing volume of relevant information. (CH)

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Analytical Potential of the Current Population Survey for Manpower and Employment Research

by 2
J. E. MORTON

December 1965

The W. E. Upjohn Inst~~itute~~ for Employment Research
300 South Westnedge Avenue
Kalamazoo, Michigan 49007

Studies in Employment and Unemployment

The Board of Trustees of the W. E. Upjohn Unemployment Trustee Corporation

HARRY E. TURBEVILLE, *Chairman*
LAWRENCE N. UPJOHN, M.D., *Honorary Chairman*
E. GIFFORD UPJOHN, M.D., *Vice Chairman*
DONALD S. GILMORE, *Vice Chairman*
D. GORDON KNAPP, *Secretary-Treasurer*
MRS. GENEVIEVE U. GILMORE
ARNO R. SCHORER
CHARLES C. GIBBONS
PRESTON S. PARISH

C. M. BENSON, *Assistant Secretary-Treasurer*

The Staff of the Institute

HAROLD C. TAYLOR, PH.D.
Director
Kalamazoo

HERBERT E. STRINER, PH.D.
Director of Program Development
Washington

Kalamazoo Office

SAMUEL V. BENNETT, M.A.
EUGENE C. McKEAN, B.A.
HAROLD T. SMITH, PH.D.
HENRY C. THOLE, M.A.

Washington Office

A. HARVEY BELITSKY, PH.D.
SAMUEL M. BURT, M.A.
HENRY E. HOLMQUIST, M.A.
SAR A. LEVITAN, PH.D.
PAUL J. MACKIN, M.A.
J. E. MORTON, PH.D.
HAROLD L. SHEPPARD, PH.D.
IRVING H. SIEGEL, PH.D.

Preface

The W. E. Upjohn Institute for Employment Research has in recent years broadened its scope and interest, moving from community-oriented research to a research program focusing on larger issues of manpower and employment analysis. In addition to the established program interests of the Institute, there has been growing concern on the part of Institute staff about the quantitative foundations upon which the researcher must base his work. On the national scene, the President's Committee To Appraise Employment and Unemployment Statistics and the two committees mentioned below also reflect concern about this general question.

It is probable that a new and lasting focus within the Institute's program will have to do with this complex and fundamentally important area which involves both content of and access to information. Although the latter is touched upon only lightly in this study, it is becoming increasingly obvious that new policies regarding access to information must be developed if the fullest potential and "return from investment" are to be realized. The scarcity of quantitatively oriented manpower economists arises, in part at least, from a lack of ready access to available data.

The primary scope of this publication is to contribute to the promotion of research in the general area of manpower and employment by surveying the supply of statistical data with a view to stimulating and facilitating their analytical use. Since the mass of relevant manpower data in this country is enormous, it was decided to approach the problem in steps and to select CPS, the Current Population Survey, as the first major data source for examination.

To achieve this objective, the present study undertakes to stimulate analytical appetites by considering not only the content of the information but also the format in which it might be usefully available for analysis; hence the emphasis on preparing the prospective data-consumer for some of the consequences of the ongoing so-called information revolution.

The new technologies which utilize electronic data-processing facilities for the storage and retrieval of information have now been widely accepted. Not well established, however, are modern techniques for the handling of statistical data; in fact, very little has been accomplished in this area, but the problem is being broached at this very time. Two committees in this country—one established by the American Statistical Association, the other by the Social Science Research Council—are exploring some of the problems involved in storage and retrieval of statistical information. A question was raised explicitly with respect to demographic data by the Statistical Commission of the United Nations during its Spring Session in 1965.

It would appear, therefore, that the time has come for preparing the subject-matter analyst and user of the data for the new information technology and its many aspects. Since the analyst needs ample time to adapt to the new data-retrieval and -manipulation techniques, it would seem appropriate to describe the storage and retrieval techniques even before they are ready to be applied and also to discuss some of their implications. The forward-looking analyst will want to include these developments in his contemplations of what method to choose in trying to answer the questions he is facing as well as in his speculations about what kinds of new problems to tackle.

A study such as the present one, because of its nature, must rely heavily on information and opinions offered by others. The author gratefully acknowledges those persons who offered help and cooperation in the development of this study; to mention all of them would seem inappropriate. Among those affiliated with federal agencies from whom advice was sought are: Gertrude Bancroft, Ewan Clague, Harold Goldstein, Morris H. Hansen, Daniel B. Levine, Rudolph C. Mendelssohn, Robert B. Pearl, Julius Shiskin, and Joseph Waksberg. It should not be inferred that these persons are necessarily in agreement with the views expressed by the author or that the Institute fully endorses his views. The author assumes full responsibility for this report.

This study was suggested by Herbert E. Striner, the Institute's Director of Program Development, who is keenly aware that successful manpower research depends upon the proper and efficient utilization of quantitative information. The author is especially indebted to him for his encouragement and help throughout the preparation of this study.

J. E. Morton

New York

September 1965

The W. E. Upjohn Institute for Employment Research

THE INSTITUTE, a privately sponsored nonprofit organization, is an activity of the W. E. Upjohn Unemployment Trustee Corporation, which was formed in 1932 to administer a fund set aside by the late Dr. W. E. Upjohn for the purpose of carrying on "research into the causes and effects of unemployment and measures for the alleviation of unemployment."

One copy of this bulletin may be obtained without charge from the Institute in Kalamazoo or from the Washington office, 1101 Seventeenth Street, N.W., Washington, D.C., 20036. Additional copies may be obtained from the Kalamazoo office at a cost of 75 cents per copy.

Contents

Preface v

I. Introduction and Synopsis 1

II. The Current Population Survey 5

III. Summary Description of Data Supply 7

IV. Statistical Information and Its Retrieval 15

V. Analytical Possibilities of the Information 19
  Data Supply and Demand 19
  Suitability of the Data for Substantive Analysis 22
  Suitability of the Data for Methodological Analysis 36

VI. Summary and Outlook 39

Appendixes

I. Administrative and Policy Considerations 43

II. Qualification of the Data 49

III. Current Population Survey Schedules and Forms 61

I. Introduction and Synopsis

One of the most prominent sources of statistical information for manpower and employment analysis is the Current Population Survey (CPS). This survey is perhaps the single most remarkable household survey ever designed: it has become a prototype for similar undertakings in this country and abroad; it has served as a vehicle for the development of much of our systematic knowledge about statistical surveys, and about the application of modern sampling theory to actual field survey operations. In addition — and most important for the present study — it has become a real gold mine of continuing quantitative information on manpower, labor force, employment, and unemployment in the United States, with hardly a rival in any other major area of comparable significance.

Chapter II points out that, in spite of its long history and availability, the analytical potential of this rapidly accumulating store of data is far from being fully exploited. Among the probable causes for this analytical underutilization are: ignorance about detail and specific aspects of the available information; comparative penury of proposed good analytical hypotheses of the kinds that require utilization of quantitative data; lack of command of proper technologies for taking advantage of the information; and, last, but certainly not least, the difficulty of gaining access to the information, especially for the nongovernment researcher, and the absence of a clearcut policy on the part of some of the government agencies involved concerning the disclosure of the information. The attitude of the several agencies toward making the information available for research purposes varies from rather positive and constructive (e.g., in parts of the Census Bureau) to negative, defensive, and protective.

The problem is not a new one. In Appendix I are outlined some of the more general considerations as they apply to the preservation of data and, especially, to the question of how to augment the research potential of the information by establishing appropriate instrumentalities and policies. After a brief discussion of the special problem of privileged information and of the particular arrangements which obtain in different agencies in order to cope with this problem, the question of making information available is being broached. The position is taken that the data-producing agency should respond favorably to bona fide research requests for information, and that it should not feel in such instances it ought to assume responsibility for the final research findings.

To illustrate some of the major efforts to come to grips with the problem, there is given a brief description of the work of the American Statistical Association's Advisory Committee on Statistical Policy, the American Statistical Association's Committee on Data Sources and Information Systems, and the Social Science Research Council's Committee on Presentation and Use of Economic Data, which recently concluded its work and submitted its final report.

Chapter III describes the data supply generated by the Current Population Survey. It gives a brief history of the survey from its beginnings in the early thirties and traces the major improvements in sample design and estimation procedure to the present. The discussion is nontechnical and is intended to draw the analyst's attention to such aspects and changes of the survey procedure that affect directly the supply of statistical data. Thus, the so-called rotation feature introduced in 1953 is described with a view to making the analyst aware of its possible use for generating information on limited cohorts for use in longitudinal studies to implement cross section analysis.

In addition to the general CPS, a monthly survey, there are within the framework of this general statistical program the so-called supplemental surveys. These surveys are in essence sets of questions added at infrequent intervals to the general CPS questionnaire, and directed to the entire CPS sample or to a subsample thereof. They are summarized here from the point of view of their potential relevance for manpower and labor force analysis.

CPS data, like much other statistical information of this kind, can be thought of as being on two different levels: (1) on micro level, i.e., information on the level of individual schedule detail; and (2) on macro level, i.e., micro data grouped together and thus forming aggregates on various levels.

The most fundamental, flexible, and hence useful information is, of course, data on the micro level. These data are the basic building stones from which all the macro data are constructed. As to CPS, practically all information on the so-called FOSDIC[1] schedule (see Appendix III) is now stored on magnetic tape by the Bureau of the Census. These microtapes also contain the weighting factors, resulting from the estimation procedure, which are needed to raise the sample results to corresponding population estimates.

---

[1] Film Optical Sensing Device for Input to Computer.

If it were not for considerations of cost and convenience, micro data would obviously be ideal because they are the most flexible and adaptable information input for further use by the analyst and researcher. However, since CPS has resulted, over the years, in considerable output of macro data, the analyst is well advised to determine first whether suitable macro data have already been produced. Within this context, attention is drawn to the experimental efforts which have been undertaken, for some time, toward a manpower and employment statistics information system which is described in greater detail in Chapter IV.

Chapter IV treats the problem of storing and retrieving statistical information in the light of ongoing developments in electronic data processing and in information technologies. Because of their potentialities, these efforts are described in some detail, so that manpower analysts may form an opinion about, and adapt their thinking to, some of the revolutionary consequences that the systematic use of high-speed electronic data retrieval and reduction might have on analytical patterns of quantitative research.

Since the output of an information system can be no better than the original input of the data, the need for qualifying CPS data arises. In Appendix II, therefore, the data are subject to a not-too-technical examination so that the analyst may form an opinion as to their limitations and qualifications for various analytical purposes.

Since CPS is a sample survey, at least two kinds of errors (sampling and nonsampling) must be considered, in addition to problems of definition, concept, and taxonomy — all of which bear on the quality of the results. Much more is known about sampling errors of estimates derived from probability samples; and since CPS is based on one of the outstanding probability designs, the analyst can form a clear picture of the effect of this kind of error. The appendix is intended primarily for the analyst who is not too familiar with modern sampling and estimations theory; it attempts to explain, in nontechnical language, what to keep in mind when forming an opinion about the quality of the estimates in the light of the presence of sampling error. Nomograms are provided to facilitate the analyst's task, and to assist him in acquiring a "feel" for the effect of sampling error, thus discouraging him from clinging to formal, numerical rules while in fact disregarding the effects of this error in the actual analysis.

Much less satisfactory is the situation with respect to the nonsampling error. Here the objective of the appendix material is mainly to familiarize the analyst with some of the ingenious attempts of the Census Bureau aimed at controlling and measuring nonsampling error, and to acquaint him with the few measures now available which he may actually use in arriving at the notion of the likely effect of such error on his analysis. An implied objective is the attempt to prepare the analyst for developments in this area which are likely to take place in the not-too-distant future so that he may take advantage of them.

A question of particular interest to the researcher is the one concerning suitability of the data for analysis and the particular analytical efforts to which they may be exposed. Even if one were limited to existing macro data, one would find that the supply of such information, including data tabulated but not published, is awesome.

So that one may look into their analytical potential, patterns of data utilization are grouped in Chapter V into two broad categories: employment of the data for substantive and for methodological analysis.

Concerning substantive analysis, the presentation proceeds by roughly considering the following styles of thought: the tendency toward disaggregation, toward dynamic patterns, toward a fresh taxonomic outlook, and toward global, overall schemes, systems, and models. Superimposed over these guidelines are the dichotomies of macro and micro data, and of cross section and longitudinal approaches. Since an exposition of this kind cannot possibly be complete, the listed cases must be taken as illustrations exemplifying the major theme, that is, the presence of as-yet-unexploited analytical potential in the data. In so doing, Chapter V points here and there to econometric and more complex quantitative applications to show the experimental potential and the ease with which it will be possible to undertake such analyses.

Specifically, analytical gaps are listed under the four headings indicated in the preceding paragraph, and ranging from analysis on low levels of aggregation, including microeconomic problems, to the analytical use of the data in the construction of descriptive, global data schemes and of more nearly theoretical models; the illustrations refer to cases involving educational attainment and levels, family and household as an analytical unit rather than the individual, and mobility problems; and they allude to the use of such analytical tools as discriminant and spectral analysis, stochastic processes, and pattern recognition.

Concerning the suitableness of the data for methodological analysis, it is felt that such data use, although perhaps not in the limelight, can be of the greatest significance because it prepares the basis for more complete and safer substantive analysis and also because it has high intrinsic value. This is particularly relevant owing to the high scientific level on which this survey program has been conducted by the staff of the Census Bureau over the years. Accordingly, the presentation of examples and illustrations here proceeds along the line of four sections: survey-technical problems; taxonomic, conceptual, and definitional questions; linking and matching of the data; and time series analysis.

The concluding chapter, Chapter VI, summarizes the general results in terms of what some of the requirements would appear to be for a fuller analytical utilization of CPS data; put differently, how a better balance may be achieved between the analytical demand and the supply in form of the CPS product. It is pointed out that data supply is not the only problem but that research-oriented demand, in turn, for empirical information is a delicate bloom which must be carefully nurtured along. Therefore, the data can do more than describe and *test* hypotheses; if properly managed, they can become a powerful catalyst in stimulating intellectual curiosity, and they could play an important role also in generating hypotheses.

Conversely, there are of course gaps in the formation, i.e., there are imbalances in the supply-demand situation because justifiable demand for CPS data exceeds the present supply. However, these situations are becoming less frequent and much less striking. Considerable possibilities would seem to exist for a stretching of the available data product by collating information from several surveys, by linking information pertaining to identical respondents over time, by matching CPS statistics with information obtained from other data sources, and the like. Ultimately, however, an expansion in the sample size — even if for fewer surveys than one every month — would seem to be unavoidable if disaggregation were permitted to proceed to analytically desirable levels.

All such attempts to manipulate and to expand the information basis must, of course, be looked at here in the light of the likely effect this would have on successful analytical efforts. Thus, it appears that the format in which the information becomes available is of signal importance, especially if the analysis is to take full advantage of the new mass-data-processing technology; and of course even more important, because crucial, is the answer to the question whether the researcher in his quest for information has the effective and understanding assistance of the agencies commanding access to the data.

II. The Current Population Survey

With the increasing acceptance in research of empirical approaches, and with the advancing statistification of the behavioral sciences, observational data, especially those of a quantitative nature, have become of the greatest importance and interest to those concerned with analysis of social phenomena. So much so have the resulting analytical procedures and methods found a home in today's research structure that it is hard to imagine how any significant findings in the area of the social sciences could have been produced without the intervention of a stage involving analysis of empirical and quantitative data. Without posing here the question as to appropriateness and efficacy of quantitative approaches, it is merely assumed that to the extent of their acceptance, the availability of the quantitative raw material has become one of the most far-reaching and crucial factors, influencing not only the choice of a particular analytical pattern and research procedure but also the selection of the specific research problem. Especially for the analysis of manpower utilization problems, these quantitative approaches are and have been of considerable importance.

The closeness of the roots of manpower analysis to demography has given manpower analysis a strongly quantitative complexion from its very beginnings. Also, manpower analysis resembles in certain respects another strongly quantitative field — economic dynamics. The early beginnings of economic dynamics can be traced to the occurrence of economic crises, i.e., to economic "pathology," long before a more nearly "physiology"-oriented business cycle analysis emerged; likewise, the important motive prompting the manpower analyst can be traced to the evil of unemployment. It is therefore no accident that the systematic collection of current statistics on manpower utilization is an offspring of the Great Depression of the thirties with its acute and far-reaching unemployment problem.

Although perhaps more nearly a historical accident, a third quantitative kinship should be mentioned here: the one between the systematic collection of current manpower utilization information and sampling survey methods. It is a historical fact of some practical interest within the present context that the early beginnings of the data collection program (Current Population Survey) were also the starting point for probability sampling as an accepted and recognized form of collecting government statistics. Much of the qualification, interpretation, and utilization of CPS manpower information must therefore be based on a thorough understanding of statistical method as it relates to probability sampling; hence, much of the appraisal of the analytical potential of the statistical data under consideration here — unlike in a more conventional situation where such data are based on complete canvasses of the census type — must be made in full awareness of the implications and effects of the partial canvass on the nature and character of the information.

Once it has been agreed upon that quantitative empirical avenues would seem appropriate for the particular analytical approach, access to and proper utilization of quantitative information become crucial issues. The importance of information cannot be judged easily. First of all, value and cost of information are two quite different properties. Considerable thought has been given to the problem of the cost of statistical data in general, and of CPS in particular. Over the past two decades or so, much effort and ingenuity has gone into the exploration of the cost situation of CPS, with a view to reducing expenditures per unit of statistical output. Improvements in the production of the information, in sampling and estimation procedures, in the control of the quality of the data, and in the streamlining of data processing are examples illustrating the high degree of proficiency with which the cost problem arising out of CPS has been handled. In spite of such efforts, any major data-production program of the government is likely to result in all but negligible cost.

First, there is the direct dollar cost paid for by taxes — by the public at large. Second, although not involving monetary expenditures directly, the collection of the statistical data often puts a heavy burden on the respondent. This burden is felt particularly where government agencies collect information from respondents in the context of other than their administrative and operating responsibilities and where, as in the mandatory survey, the government claims a right to exact the information. In spite of the fact that CPS is not a mandatory survey, the production of such statistical data tends to exhaust the information source which lends CPS data the character of a national resource; CPS data should, therefore, be carefully husbanded in the same manner as any other high-cost resource, and their analytical value should be optimized.

It would thus be tempting to subject CPS statistics to a cost-benefit type of calculus. Such a calculus, which would allocate to particular cost-incurring inputs the benefits or values to be derived from the corresponding outputs, is difficult in any complex situation.[1] For statistical production systems and their uses, such attempts have as yet hardly been undertaken. Unlike costs of production of the statistical data, the results of efforts to find a solution to the question of valuation have been tenuous, to say the least.

First of all, the conceptual aspects are here more complex, and no satisfactory models corresponding to cost-evaluation patterns are as yet available. It stands to reason that information *per se* has no value, whatever its cost, other than the value derived from its use. Since the interest here lies in the value of CPS-type information for use in the analysis of manpower utilization, the evaluation involves analytical uses of the information rather than administrative or operations-oriented uses. Hence, it would seem appropriate to gauge the analytical value of CPS-type manpower information by its actual and potential analytical utility. Put differently, an attempt to increase the value of CPS information should explore not only promising and costwise feasible modifications and expansions of the statistical product but also the broadening of the analytical use of existing data. It would seem that much less has been done in the direction of increasing the analytical utilization of the data than in optimizing the cost of producing them.

Part of the problem may well be the setting within which general-purpose, mass-data-collection systems of the government are initiated and propagated — the division of labor between the data-gatherer and the data-user and the ensuing difficulties of communication.

Be that as it may, it would seem at first sight that the analytical use of CPS-type information might profitably be expanded, if for no other reason than because of the enormous importance and analytical challenge presented by the manpower problem, a long-forgotten stepchild of the behavioral sciences. Also, it appears that there might be more information in existence than is now actually being subjected to analytically oriented exploitation.

What, then, might be some of the reasons for the less than full exploitation of the analytical potential of the data? The following factors stand out among the many other ones that, altogether, seem to contribute to the analytical lag which, of course, is a general phenomenon characteristic of nearly any major system of general-purpose statistics:

1. Lack of thorough knowledge about the existence of information collected within the framework of CPS and the limited usefulness, especially for more complex kinds of analysis, of the data in the form in which they are published and generally made available.
2. Inaccessibility of certain kinds of information to nongovernment analysis because of administrative practices and policies, including the thorny problem of disclosure of confidential information.
3. Limitations in the availability to analysts of data-processing facilities, including not only the so-called hardware — computing equipment — but also the software — the programs, procedures, and instructions which, couched in the language of the computer, are needed to communicate with the equipment in order to take advantage of it.
4. And closely related to these three factors, a relative penury of hypotheses and of rewarding problems of the kinds that would require analysis of the quantitative information generated by CPS. In other words, there seems to be a shortage of sound analytical questions involving, where appropriate, model-building and related quantitative methods and patterns of research.

The first limiting factor, the relative ignorance about the existence of the information coupled with the frequent lack of suitable format, is treated in Chapter IV as an obstacle in the way of physical access to the data. The general situation with respect to point (2), i.e., the administrative obstacles in the way to fuller analytical exploitation of the data, is sketched in Appendix I on administrative and policy considerations. As to points (3) and (4), illustrations are given in Chapter V of potential uses of the data in the hope that they will stimulate the analytical appetite of the prospective data-consumer while keeping him aware of the nature and limitations of the statistical end product. In particular, Chapter V attempts to lead the future data-requester to the point of view that the CPS data-production, -storage, and -retrieval functions should be regarded as parts of an integral system, and to show that this viewpoint might in the not-too-distant future influence the formation of hypotheses and the very art of "scientific discovery and creation"[2] in the field of manpower and employment research.

In summary: It would appear that, as for other types of information, there seems to be an unused analytical potential contained in CPS-based manpower data. One way, therefore, of increasing the analytical value of the information is to reduce the gaps, flaws, obstacles, and limitations which may account for the lag in analytically oriented demand for information. In the subsequent discussion, special attention is given to the identification of some recent developments and circumstances which tend to remove such hurdles and which therefore tend to enhance the analytical value of the data and stimulate research into problems for whose solution CPS-type manpower information is an invaluable source of empirical raw material.

---

[1] See, e.g., Charles J. Hitch and Roland N. McKean, *The Economics of Defense in the Nuclear Age*, Cambridge, Harvard University Press, 1961.

[2] A. Koestler, *The Act of Creation*, New York, Macmillan, 1964.

III. Summary Description of Data Supply

The data supply of primary interest here is that generated by the Current Population Survey (CPS) and bearing on manpower and labor force characteristics. This survey is one of the most distinguished and oldest parts of the federal government's survey program. Its beginning can be traced back to the late thirties when, during the Great Depression, the problem of estimating unemployment became particularly acute. It was during the latter part of this decade that the Works Progress Administration undertook an inquiry into the problem of measuring unemployment,[1] which resulted in the sample survey of unemployment, a monthly survey of unemployment undertaken by the Works Progress Administration beginning early in 1940.[2] In August 1942 the survey operations were moved from the Works Progress Administration to the Bureau of the Census. The survey was substantially revised in October 1943, when it became, more or less, the sample design as we know it today. Hence, for the practical purposes of the analyst, it may be assumed that the end of 1943 is the starting point for the information on manpower and labor force generated by the Current Population Survey. This is to say that at that time the survey sample was put on a modern probability basis. Over time, it was substantially improved and expanded; it became in essence the type of survey which we know now under the name CPS.

Subsequently, changes in the design were undertaken so as to reduce successively the sampling variances; these reductions were achieved through improvements in the estimation procedure, through expansion of the sample, and through various changes in the survey procedure. The major steps in this development of direct importance to the analytical user and interpreter of the data have been the following:[3] After the initial design of the 1943 sample, which modified substantially the sample taken over from the Works Progress Administration, and which was based on a sample of 68 primary sampling units, the first major revision was undertaken in July 1945 when the four basic employment status items were introduced into the questionnaire; they are still being used today. The wording of the questionnaire as introduced at that time defined labor force so as to include also part-time and intermittent workers, in particular unpaid family workers working at least 15 hours per week. Since most of these persons had been excluded from the labor force under the questionnaire wording prior to 1945, the new wording resulted in a noticeable increase in the estimate of labor force; also, from then on all those were included in the labor force who worked for pay or profit as little as one hour a week.

Since September 1953, the explicit introduction of "color" as a term in the ratio estimate[4] made it possible to arrive at estimates in absolute figures for nonwhite persons, whereas previously only percentage distributions for whites and nonwhites were published. Earlier that year, a considerable improvement in the estimates was achieved by introducing 1950 population census figures into the computation of ratio estimates.

Beginning with May 1955, questions were added to the monthly schedule[5] on reasons for working less than 35 hours a week, and on the definition given to part-time work — questions which prior to the spring of 1955 were asked only four times per year or less often.

Minor changes, helpful if national totals were to be explored but possibly disrupting to smooth continuity of series if figures were to be used for only small special subgroups, were introduced in January 1957. "Persons on layoff with definite instructions to return to work" and "persons waiting to start new wage and salary jobs within 30 days of interview" were put into the "unemployed" category; prior to that date they were counted as employed.[6]

The above-mentioned modifications and additions are of definitional and conceptual nature, and of interest and importance primarily where a study over a period of time is intended or implied, such as in time series analysis.

---

[1] Lester R. Frankel and J. Stevens Stock, "On the Sample Survey of Unemployment," *Journal of the American Statistical Association,* Vol. 37 (1942), pp. 77-80.

[2] For historical material and references of interest to the analytical user of the information, see especially U.S. Bureau of the Census, *The Current Population Survey: A Report on Methodology,* Technical Paper No. 7, 1963, pp. 1 ff.; U.S. President's Committee To Appraise Employment and Unemployment Statistics, *Measuring Employment and Unemployment,* 1962, pp. 30 ff.; and the publications mentioned in footnote 3 below.

[3] The most recent summary description is given in U.S. Bureau of Labor Statistics Report No. 279 — U.S. Bureau of the Census, *Current Population Reports,* Series P-23, No. 13, "Concepts and Methods Used in Household Statistics on Employment and Unemployment from the Current Population Survey," June 1964, pp. 14 ff.

[4] For a description of this part of the estimating procedure, the so-called second stage, see Census, *The Current Population Survey: A Report on Methodology, op. cit.,* p. 54.

[5] The present questions 21 A, B, and C.

[6] Those "in school during the survey week and waiting to start new jobs" were moved outside the labor force altogether into the not-in-labor-force group. The latter shift has only minor statistical implication since the number involved is small — unless the analytical problem is one for which these persons are of particular importance.

Different kinds of modifications are those affecting sample size, sample design, estimation procedure, and the like — in other words, technical improvements in the survey. Mentioning only those major ones that may affect the analytical potential of the data, there was introduced, in July 1953, a modification of the system of rotating the sample with a view to obtaining some year-to-year overlap. Prior to the summer of 1953, households were interviewed once per month for six months and thereafter an entirely new group of households was selected to be interviewed, in turn, for another six months, etc. The objective of this rotation plan was, quite obviously, to avoid exhausting the response willingness in the sample households, and other attritions and disturbances usually connected with a permanent panel. In 1953, a system of partial rotation was adopted which is still in operation today. There are at least two aspects to this partial rotation system which should be of interest to the analyst; therefore, the operation will be described in somewhat greater detail than would otherwise be warranted.

In essence, the system adopted in 1953 results in a given set of households being in the sample for four consecutive months; then they are removed and after a rest period of eight months they are again included for four more consecutive months. The system is therefore occasionally referred to as a four-eight-four rotation system. Each household is in the sample for eight months, and once the system is underway one-eighth of the sample of households must be replaced every month by new households not previously in the sample. This monthly replacement is referred to as a rotation group.

From the point of view of the analyst who is concerned with longitudinal rather than cross section patterns, and who wishes to take a cohort type of approach, this rotation system provides a good deal of relevant material although not as much as the work-history type of data would provide. On the other hand, the kind of information provided by the Current Population Survey is, for many purposes, considerably more rewarding than that provided by the work-history data which are maintained on a broad basis by the Social Security Administration in its continuing work-history sample.

Looked at from this viewpoint, the four-eight-four rotation system is of interest to the analyst because it produces at least some information on sample households overlapping over time — that is, from interviews for the same households over successive months.

In particular, the rotation pattern results in limited cohorts — limited because of their selective exposure to interview over time; as the system now works, the following chains of interview exposure are being produced: pairs, triplets, and quadruplets of successive months; single month, three-quarters of a year, later quadruplets of successive months, and in a similar pattern — always eight months apart — pairs and triplets, triplets and pairs, and quadruplets with a single month following after eight months. Put differently, cohorts can now be observed over maximum periods of one-and-a-half years, containing a "sliding" blank interval of eight months.

These several exposure combinations will, of course, happen for different proportions of the sample. For instance, households interviewed over pairs of neighboring months will amount to three out of four households in the entire sample; the combination consisting of a household being interviewed once and, after eight months, for another four consecutive months as well as the symmetrically opposite pattern of four consecutive monthly interviews followed, after eight months, by just one more interview will apply to only one out of eight of the sample households at any given time.

Another property of these patterns is the presence in the sample of about one-half of all households in a given month as well as in that month a year earlier; put differently, there is implied in the system a 50 percent year-to-year overlap of the sampled households.

In addition to the obvious effect of the introduction of the partial rotation system on the analytical potential of the data — analytical in terms of substantive coverage — this rotation system has made possible another improvement in the direction of quality rather than substantive coverage of the information: it has improved the way for the adoption of a final phase in the estimation procedure which is referred to as the composite estimate, to be discussed briefly below.

In February 1954, the original 1943 version of the CPS sample was expanded from 68 primary sampling units to 230 sample areas, while retaining the overall size of the sample — about 21,000 households. This expansion, together with the introduction of the "composite estimate,"[7] achieved a considerable reduction in the sampling variance equivalent to an approximate doubling in the size of the sample for many of the relevant characteristics.

In May 1956 the CPS sample was further enlarged, from 230 areas to 330 primary sampling units. This time, however, the number of sample households was also increased. The expansion amounted to an increase from 21,000 sample households, in the sample up to that time, to 35,000 households beginning with May

---

[7] This estimate is, in principle, a weighted average of the estimate for the current month arrived at directly, and a corresponding estimate arrived at indirectly by utilizing the overlapping part of the sample resulting from the partial rotation; this indirect estimate arrives at a monthly figure by adding to a given month the estimated change from that month to the next. The procedure is described in Census, *The Current Population Survey: A Report on Methodology, op. cit.*, pp. 54 ff.

1956. This enlargement of the sample is estimated to have improved the reliability for salient estimates by about one-fifth.

The latest revision of the sample was completed in March 1963. The new sample, which incorporates a series of revisions and technical improvements,[8] consists of 357 primary sampling units, that is, 27 sampling areas more than its predecessor. Although the number of households (the ultimate sampling units) in the sample remained at the level of about 35,000 per month, the increase in the number of primary sampling units together with the other improvements in design achieved a further rise in precision of the estimates.

So far, the discussion has been in terms of the Current Population Survey *per se*. In addition to the supply of data provided by this general CPS — a monthly survey which has produced much of the basic information underlying the Bureau of Labor Statistics' monthly report on the labor force — there are the special, intermittent inquiries which are directed to the same CPS sample or a subsample thereof at irregular intervals or regularly, but much less frequently than once every month. These supplemental surveys furnish much of the basic information described in the Special Labor Force Report of the Bureau of Labor Statistics,[9] and they contain much statistical information with considerable analytical potential.

These special surveys are in the form of supplemental questions added to the regular CPS form and used for interviewing the sample households or a subsample of the CPS households.

The ones more immediately relevant for manpower analysis are:[10]

1. The February Supplement on "Work Experience."
2. The March Supplement on "Income, Educational Attainment, Family Characteristics, etc.," also referred to as the Microcensus.
3. The May Supplement on "Dual Jobholders."
4. The October Supplement on "School Enrollment."

Information produced by supplemental questions and by *ad hoc* surveys directed to the CPS sample of households is not in all instances labor force oriented; even so, it may frequently be of considerable interest to the manpower analyst inasmuch as it may be possible to collate the information with other, more directly manpower-oriented data derived from the same sample of households. Therefore, not only the four listed supplements but also other inquiries directed to the CPS sample, although they may have quite different objectives, may contain information, the interpretation and analysis of which would facilitate an understanding of manpower utilization patterns.

The four above-mentioned surveys are conducted annually, and all but the items in the March Supplement are addressed to the entire CPS sample; the income question of the Microcensus is asked of a 75 percent subsample selected from the total CPS sample. All four surveys go back as far as 1959 or farther; with the exception of the Dual Jobholders survey, they were conducted during the same month every year. In other words, these surveys have been conducted during the entire period for which microtapes are available.[11] Where published information is satisfactory by itself, the tabulated data can, of course, be traced farther into the past; but it is not possible to retabulate the data. Therefore, the data have to be used in whatever form the historical tabulations have been undertaken; this also means that they have to be used on the particular level of aggregation and detail on which the published information was tabulated.

In addition to these four surveys, which are sponsored by the Bureau of Labor Statistics, or jointly by the Bureau of the Census and the Bureau of Labor Statistics, there are other recent but more nearly intermittent and irregular surveys of importance to manpower analysts which are sponsored by the Bureau of Labor Statistics or by other agencies within the Department of Labor. Among them might be mentioned:

> A survey on premium pay which provides information on the number of wage and salary workers receiving overtime pay; and a survey of labor force attachment of women to determine reasons for their entering or leaving the labor market. The latter survey covers a subsample of 2,000 women, selected from among recent entrants into and withdrawals from the labor force recorded in the February 1964 CPS sample.
>
> A survey on the training of workers 20 to 64 years of age, with the objective of establishing information on the kind of vocational training they had had in schools, in apprenticeship programs, or otherwise. This survey, which was sponsored by the Office of Manpower, Automation, and Training, was based on a CPS subsample of about 25,000 self-enumerating persons.
>
> A survey of the employment status of migrants, based on answers in the CPS sample from persons who had changed their county of residence since the preceding year. The questions were aimed at reconstructing employment status of the movers before their move, and the reasons for the move.

---

[8] *Ibid.*, Supplement II.

[9] Prior to July 1959, the labor force information contained in CPS was published by the Bureau of the Census in its *Current Population Reports*, Series P-57. The Special Labor Force Reports are reprints of articles published in the *Monthly Labor Review*, with additional statistical material and explanatory notes; they have been published since February 1960.

[10] See Appendix III.

[11] In general, microtapes were not available prior to 1959.

The out-of-school-youth survey, a combination of self-enumeration and interview of a CPS subsample of about 2,000 persons from 16 to 21 years old. The survey questions were designed to throw light on the educational background of the sampled young people, their work experience, and their general economic and social characteristics.

A duration-of-employment survey was undertaken, including additional questions directed to the employed persons in the CPS sample with respect to the length they held their current jobs.

A detailed survey of persons in the CPS sample who had experienced substantial unemployment, in order to obtain information on their work history and general background.

In addition to the surveys which were sponsored or cosponsored by the Department of Labor, there were surveys sponsored by agencies other than the Department of Labor; these were in the form of supplemental questions to the CPS sample of households or a subsample of it. Some of these may be of interest in the analysis of the economics of manpower utilization because they permit, at least in principle, a collating of supplemental information with the basic labor force and general demographic information on the sampled individuals. Among such surveys, conducted during the past five years or so are:

The annual supplements to the CPS on hired farmwork, sponsored by the Department of Agriculture. These inquiries are directed to the full CPS sample in order to gather information on the number and on the economic and social characteristics of migratory and other farmworkers.

A survey early in 1963 of senior citizens directed to a subsample of about 8,000 married couples and individuals 62 years of age or older, in order to obtain for the Social Security Administration, which sponsored the survey, information on a large array of socioeconomic characteristics including employment background, income, living arrangements, etc.

At the end of 1962, surveys of veterans in the CPS sample were sponsored by the Veterans Administration in order to ascertain, in addition to socioeconomic characteristics, the use that veterans had made of the various special programs, including training programs. A related survey was conducted, in January 1960, of veterans' widows.

In March and April 1962 a self-enumeration survey of a CPS subsample of about 30,000 respondents was undertaken for the University of Chicago (via a research grant from the National Science Foundation) on intergeneration occupational changes. Information was obtained on the principal occupations of the fathers of adult CPS interviewees in order to reconstruct occupational changes over that generation as well as to interpret the socioeconomic patterns associated with occupational mobility.

In November 1961 supplemental questions were added to the CPS schedule on pay scales and working arrangements for domestic servants, babysitters, and other private household workers. This survey was sponsored by the Social Security Administration with the primary objective of assessing social security coverage.

Finally, the so-called quarterly surveys of intentions, originated under Federal Reserve Board sponsorship, should perhaps be mentioned because of the wealth of information they contain on purchase intentions as well as on actual purchases. If such information could be correlated with pertinent labor force and demographic information, there would be exploratory possibilities where, e.g., the interrelation between labor force status and consumer demand is of analytical interest.

Next to scope and substantive and time coverage of the statistical data, the question of their availability is of decisive importance to the prospective analyst. Availability, therefore, is here not interpreted in the general sense of existence of data but qualified so as to imply reasonable access to the information. On the other hand, the term, "availability," is here used without prejudice toward a particular government agency's decision as to whether, and to whom, such information should actually be supplied. In other words, information will not be considered that might become available only after substantial historical research and exploration of sources to which access would seem to be extremely difficult and probably not feasible at all in most instances.

With this qualification in mind, it is useful to think of the CPS-type information as being on at least two different levels:

1. The statistical raw material contained on the personal record of members of individual households. This is information on the level of greatest detail; it is also information on the lowest level of aggregation, and in this sense the information consists of micro data.
2. Estimates resulting from the condensation and reduction of the micro data in various ways and on various levels of aggregation ranging from relatively low ones to the very highest.

For purposes of data use by the analyst, the macro data which underlie this kind of information may, in turn, be subdivided into (a) macro data reaching the state of publication, and (b) all other macro data — that is, unpublished data which accrue during the various stages of data processing. This latter type of information may have been recorded on punched cards, on magnetic tape, or in some other form such as on worksheets, in posting books, and the like.

Concerning Current Population Survey data, availability in an operationally feasible shape is closely tied to the particular form of data processing in use by the Bureau of the Census at any given time.

Microtapes, that is, records of data on the level of the original detail contained in the personal records resulting from individual interviews, are generally readily available for the period beginning with 1959 when Univac 1105 was put into operation. This information is collected through monthly interviews of sample households, and it refers to the civilian (noninstitutional) population, 14 years of age and older.

On these microtapes there is recorded nearly all the information in the detail in which it was collected and recorded on the CPS form.[12] Where open end questions are involved in the several parts of Item 26 of the FOSDIC schedule, the coding is undertaken on the basis of the Standard Industrial Classification and the three-digit code according to the alphabetical index of occupations and industries.[13]

Strictly speaking, however, the data recorded on microtape are not the original raw data as recorded on the CPS forms in a literal sense; they should rather be thought of in terms of edited raw data which, where appropriate, have already been edited and modified. The information on the microtape is therefore much more useful than the original schedules, and saves the researcher time and energy.

From the prospective analytical user's point of view, the tape can be said to contain the following sections of information: the just-mentioned, edited raw data, including imputations and the like; the so-called re-codes which do not contain any new information but are useful for subsequent tabulations and reduction of the information by grouping the data into various taxonomic groups and subgroups; and the basic weights including sampling fraction, nonresponse adjustment, and the weights corresponding to the several stages of ratio estimation intended to reduce the variance between primary sampling units and to adjust the sample estimates to independent estimates of the revelant population groups and subgroups derived from carrying forward latest population census data, month by month. Not recorded on the tape are the so-called composite estimates which require the averaging of two sets of data (see page 8); this task is performed directly by the computer which prints out the results for specific tabulations without, in turn, recording these results themselves on the microtape. In actual analytical usage of the microtape, the composite estimation procedure could, of course, be programmed into the data-processing operations; or other forms of composing estimates could be undertaken, for instance, by averaging in various ways and directions, including not only preceding but also following information. Finally, it should be pointed out that omitting the composite estimation stage altogether would, of course, not bias the result but would increase the sampling error, especially where month-to-month changes are to be analyzed.[14]

This information, which altogether refers to about 70,000 individual records based on completed interviews of approximately 32,000 households, results in four magnetic tapes, the so-called F-reels, one set of four for every month going back to 1959.

As will be remembered, the sample which yields this micro information is the 330-area sample which was introduced in 1956; the only change therefore in sample size, hence in sampling error, which the user of microtapes need consider is the expansion of the sample in 1963 from 330 to 357 primary areas; this, however, leaves unchanged the total number of approximately 35,000 households to be interviewed every month. This increase, together with some other adjustments in the sample design, resulted in a reduction of the sampling error, as is briefly pointed out in the sections on sampling variability of the data.

As to geographic coverage, the prospective analyst may be able to benefit from the fact that 263 primary areas in the new 357-area sample were simply taken over from the earlier sample; 53 primary areas were entirely new and the balance of 41 areas were also carried over from the earlier sample but only after some changes in delineation of the primary sampling units, which in the particular instance may have resulted in adding, subtracting, or adding as well as subtracting some area to or from the primary sampling unit as defined for the 1956 sample. Quite apart, therefore, from considerations of sample size and sampling variability, a comparison over time of estimates on lower than national levels of aggregation might have to consider the possible changes in definition of area where the level of aggregation is sufficiently low to be influenced by such a change.

With this general restriction in mind, it would seem that the stored microtapes contain reasonably comparable information over the years from 1959 to date.

As to micro information derived from the several supplemental population surveys, the situation is similar since the sample on which these surveys are based is either identical with that for the general Current Population Survey or is a subsample of that sample. In passing, it might be mentioned that the chance that microtapes are going to be preserved is considerably better for supplemental surveys than for the general Current Population Survey. First of all, much less information — only answers to a few supplemental questions — is involved. Furthermore, the supplemental surveys are by and large annual surveys; for this reason alone the storage problem is reduced to less than one-tenth of that for the general Current Population Survey.

---

[12] See Appendix III for Form CPS-1, the so-called FOSDIC schedule.

[13] U.S. Office of Statistical Standards, *The Standard Industrial Classification Manual,* rev. 1957, and 1963 Supplement; and the *Classified Index of Occupations and Industries,* 1960.

[14] The increase in sampling variance resulting from omitting the stage of composite estimation would be of the approximate order of 10 percent.

Concerning the time period for which other than published information might be available, it should be noted that for the period from 1954 to 1959 (the period during which the Census Bureau used a Univac I installation), it might be possible to reconstruct some of the micro data. It is doubtful, however, that anything of analytical value can be readily reconstructed for any year prior to 1954, short of a major historiographic effort and much scrutiny. As might be expected, the situation is more complex where macro data are involved. First of all, a great number — indeed an infinite number — of different levels of aggregation are conceivable along the continua of many different dimensions: e.g., aggregation along geographic, functional, demographic, or labor force status taxonomic dimensions. The problem may be further complicated by the need for producing separate estimates for different classifications and cross-classifications of the original data. Finally, a great variety of tabulations linking individuals through time is conceivable as exemplified by the so-called gross change measures.

In view of the enormity of analytical possibilities and the ensuing potential demand for aggregated data, it is convenient to divide this macro information into two parts: on the one hand, macro statistics already produced — though perhaps for purposes quite different from those intended by the particular analyst who would like to avail himself of the information; on the other hand, the rest of conceivable and practicable macro information which could in principle be furnished if administratively and financially feasible.

Macro tabulations are often produced within given publications programs and with a definite view to making the resulting tables available in printed or otherwise reproduced form. Readily available to the analyst are the usual monthly publications in which the Current Population Survey type of manpower information appears. Examples are the *Monthly Report on the Labor Force,* and *Employment and Earnings,* both publications of the Bureau of Labor Statistics.[15] In addition, annual averages are published — since 1959, as a rule, in the detailed tables appended to the Special Labor Force Reports.

The researcher may be even more interested in tabulations which, although undertaken to the point of print-out, have never been released for general publication for various reasons.[16] With the expanding use of high-speed electronic data-processing equipment, the amount of tabulated aggregative material of this kind has increased rapidly and is assuming awesome dimensions, frequently exceeding even the capacity for inventorizing, coding, and storing the information for subsequent retrieval. Thus such tabulations have, in fact, assumed a character of *ad hoc* information, extremely difficult if at all possible to locate or reconstruct.

As a general rule of thumb, it may be assumed that the unpublished tabulations usually are on lower levels of aggregation, and also otherwise in greater detail than the tabulations that find their way into the published tables and reports. It is therefore of some concern, if analytical exploitation of the unpublished macro data is being considered, that the macrotapes, the so-called monthly final data tapes, usually have not been preserved once print-out of the tabulated information has been completed. However, the print-outs themselves are available, primarily from 1954 to date; that is, for the period during which first, Univac I and later, Univac 1105 were in operation. Prior to 1954 a good deal of macro information had been preserved in the form of so-called posting books. The transfer, in July 1959, of the responsibility for publication of labor force estimates from the Census Bureau to the Bureau of Labor Statistics, while the Census Bureau continued to collect the information as agent for the Bureau of Labor Statistics, also resulted in a transfer of posting books and macrotape print-outs to the Department of Labor. Recent developments in the Bureau of Labor Statistics are not unlikely to result in a manpower data bank which should be of the greatest interest to the potential analytical user of the macro information.

With respect to the data from supplemental surveys, that is, the supplemental questions added to CPS forms from time to time, the picture is considerably more involved since such supplemental questions are usually added by the Census Bureau at the request of another agency; the sponsoring agency, which as a rule receives the information, may preserve it or not, depending on the circumstances. As already noted, supplemental questions of this nature will usually be of importance to the manpower analyst only if dovetailed with other information from the CPS schedule, mainly with data directly pertinent to the labor force status and to related manpower characteristics of the respondent. Micro information will therefore be needed as a rule, unless already published macro information can be used.

In summary, it can be said that the supply of data here of concern is derived from two sources: the main-stream of CPS data accruing on a monthly basis; and a system of large and small tributaries, irregular in nature, including the supplemental annual surveys, and a host of intermittent and *ad hoc* minor surveys — all of them directed to the CPS sample of households or a subsample thereof. In following this stream of information to its origin, one notes a remarkable stability over time in the general CPS framework; it is this relative stability which, among other properties, makes the CPS information such valuable material for longitudinal and time series analysis.

---

[15] Prior to the transfer, in July 1959, of this function from the Bureau of the Census to the Bureau of Labor Statistics, the data were published by the Bureau of the Census in *Current Population Reports,* Series P-57.

[16] In addition to the usual financial and administrative restrictions of agency publication programs, the reasons for nonpublication include the tentative and experimental character of some tabulations, the relative size of the sampling error for levels of detail on which the data may have been tabulated but not published, lack of justifiable and broad public interest in the particular information, and the like.

As to the present availability of earlier data, much depends on whether the analysis can be usefully undertaken with data already published or, at any rate, tabulated. Where recourse must be had to the original micro data derived from the schedules themselves, it is unlikely that much information can be readily retrieved for the years preceding 1959.

IV. Statistical Information and Its Retrieval

Extraordinary developments in various fields of human activity have characterized the second half of this century; they have been referred to variously as explosions, revolutions, and the like. Among them is the imposing rise in the volume of information, characterized sometimes as an information explosion or revolution. Concurrently with this sustained and abundant enlargement of our store of knowledge, there has been observed a growing difficulty in bringing together supply of and demand for the information. The problem — in essence one of communication — has been threatening to overshadow the production of the information, and considerable effort has been and is being directed to the several facets of the problem. Most of the attempts have been undertaken with a view to communication of verbal information, and very little has been learned in this context about quantitative information of a statistical nature — a situation particularly remarkable and surprising in view of the close relationship between statistics and electronic data processing.

In the field of manpower statistics, as in many others, there exists no information catalog or current data inventory. The time-honored handbooks and introductions to the use of statistical data[1] are not designed to be substitutes for a data inventory. Neither are they current between the widely spaced revisions. The only source now available from which an opinion can be formed about some of the detailed aspects of the current manpower data supply are the tablestubs and headings in publications such as *Employment and Earnings,* published monthly by the Office of Manpower and Employment Statistics of the Bureau of Labor Statistics; the Special Labor Force Reports, which are reprints of articles in the *Monthly Labor Review,* and which generally contain additional detailed tables and explanatory notes; *Labor Force and Employment,* which lists annual averages for many of the series tabulated on a monthly basis; and the Statistical Appendix to *The Manpower Report of the President,* which gives annual data. But these and similar sources can be used only to identify the data output and to see the level of detail at the time of publication. They do not reflect statistical production, planned or underway, and ordinarily they cannot be used as a source list for the great volume of statistical data which are never published.

It thus appears that practically all the attempts to acquaint the data-user with the complex world of statistical information are in the nature of general guides; they tend to describe past situations, and they give, intermittently, a picture of the major outlines of established data publication patterns of completed statistical activities. They are intended to serve as a first introduction rather than as an operational index and detailed inventory for the data-requester who desires statistical information for analysis. Hence, it is now mostly up to the data-user to become his own expert on statistical inputs as well as on the intended analytical output.

In actual research practice, the analytical potential of a given body of information depends not only on its content and quality but also, and to a considerable degree, on the format in which the data confront the analyst. Things other than the information inventory itself are of importance to the analyst: a good deal hinges on the general character and on the technical aspects of the information layout, on the indexing and coding of the material, and on the way in which the technical retrieval problem is solved. For instance, a question of considerable practical importance to the analyst is whether data are available in printed publications, or in some other format.

To date, the analytical user of CPS data has been limited, as a rule, to statistical data which were published in one way or another, or which had become available direct from machine runs for use as worksheet material and the like. Because of clerical limitations, priority on the use of information of this kind has generally, although not without exception, been given to staffs of government agencies. Concurrently with this traditional form of communicating statistical data to the analytical user, efforts have been underway for some time to explore new avenues to the problem of information dissemination. Most attempts of this nature have been known under the name of information storage and retrieval.[2] Although by now there has been developed a good deal of thinking and writing on this subject, and although the last years have witnessed an impressive growth of the new discipline of the so-called information sciences, most of these efforts have been devoted to information of a textual nature as illustrated by literature abstracts, scientific papers, and the like.

---

[1] For instance, Philip M. Hauser and William R. Leonard, *Government Statistics for Business Use,* 2d ed., New York, Wiley, 1956, especially Chapter 13 on Labor by Charles D. Stewart; Laurence F. Schmeckebier, *Government Publications and Their Use,* rev. ed., Washington, Brookings Institution, 1961; U.S. Office of Statistical Standards, *Statistical Services of the United States Government,* rev. ed., 1963.

[2] This terminology, which usually refers to documentary rather than statistical information, has been in use for about 10 years and originated within the general context of library services. See, e.g., J. Becker and R. M. Hayes, *Information Storage and Retrieval,* New York, Wiley, 1963.

It may therefore be considered a coincidence of particular note that it is in the very area of manpower data that a major step forward may be undertaken toward developing a statistical data-oriented information-storage system, adapted to modern electronic data-processing technology. For some time, the Data Systems Research and Development Unit in the Office of Manpower and Employment Statistics of the Bureau of Labor Statistics has engaged in imaginative planning and experimental work which may eventually lead to a statistical data-storage and -retrieval system whose possibilities, once fully operative, would encompass the entire area of data manipulation for use by the prospective data analyst.

The prospective analyst should realize that, whereas it is appropriate to think in terms of only one micro level (in the present instance, the level of an interviewee's individual record), there are a great many macro levels. Put differently, there is an infinity of ways of aggregating data pertaining to individuals: by geography, by demographic groups, by labor-force-status categories, by functional groupings such as industry and occupational attachment, and by many other taxonomic classes. The aggregation may proceed along such taxonomic lines singly, one at a time, or it may involve simultaneously two or more. The resulting number of possible cells is awesome, and the relative ease with which a modern computer can perform the task should not deceive the analyst about the need for a great deal of judgment when it comes to appraising and evaluating the strength and carrying power of the data.

To convey a general idea of such a system's capabilities as viewed through the eyes of a prospective data analyst, a brief summary is given of some of its characteristics in an extremely simplified manner: The input (statistical data) is first entered in such a system, then updated, retrieved upon request, subjected to statistical reduction according to instruction, and finally printed out by the computer in the desired format. The following separate and distinct phases must therefore be accommodated by the system:

1. Information entry into storage.
2. Updating of the stored information.
3. Retrieval of the stored and updated information.
4. Statistical manipulation and reduction of the data.
5. Print-out of the requested information in the stipulated form.

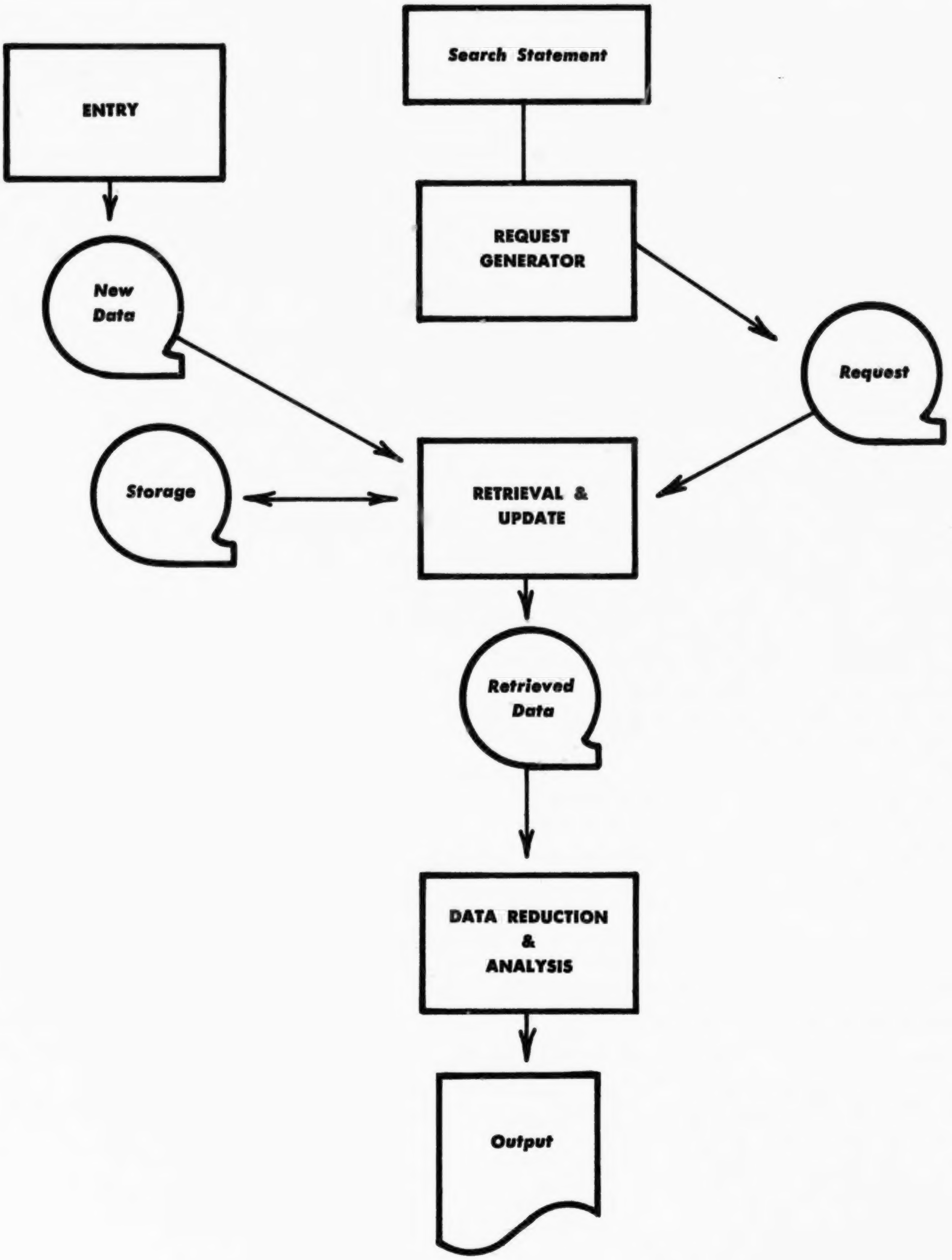
For practical purposes of the prospective analytical user of the data, it would seem quite adequate to think of such a system with all its implications for electronic data processing and programming, for the necessary coding and indexing efforts which must precede the very "inputting" of the information, and for the simple as well as very elaborate statistical, econometric, and similar data manipulations and computations in terms of a schematic pattern as represented by, say, a flowchart. Such a flowchart, in spite of its limitations, does convey the salient features of the system more succinctly than any not-too-technical textual presentation would (see figure on page 17). Correspondingly, in interpreting the set of flows described by the diagram, the prospective data-requester may visualize this set in terms of four subsets or segments of the flowchart, matching the four parts of the system.

1. *Entry segment.* Here the CPS data would first be rearranged in a form suitable for storage in the system. This requires, among other things, a sufficiently detailed and comprehensive taxonomy in the form of a generalized code. It would be one of the first and foremost tasks of a completed system, and — as already mentioned — one of considerable importance to the prospective data-user, to generate its own table of contents. This table of contents would be a systematic presentation of all possible effective code combinations — effective in the sense that the resulting data cells would not be empty.
2. *Maintenance segment.* Here the updating of the series in storage would be undertaken as new information becomes available, month after month. It is in this segment of the system that replacement, modification, and addition to the items in storage take place. It is also here that data are again expelled, so to speak, from storage in answer to a search request through which retrieval of the specific data is initiated.
3. *Request-generator segment.* In order to activate the retrieval operation, a search question must first be presented to the system. For this purpose a request-generator segment must be provided. After translation into appropriate machine language by this request-generator, the search question submitted by the analyst is passed on to the appropriate parts of the system, including specification of desired data manipulations, computations, and layout.
4. *Data-reduction segment.* A final segment must be provided to perform the all-important task of data reduction, that is, the statistical and mathematical manipulation of the retrieved information or the layout of tables.

In spite of the very impressive capabilities of the retrieval operations, it is the data-reduction potential that would convey to such a system a near-science-fiction character. To appreciate fully this potential, it must be realized that the computer, in spite of its undeniable capacity for storage and retrieval tasks, is a computer first of all; that is, it is a very efficient and a very fast data-processing device. Rather, therefore, than thinking of a system like the one just described as a storage and

Flowchart

Data-Storage and -Retrieval System


Search Statement
ENTRY
REQUEST GENERATOR
New Data
Request
Storage
RETRIEVAL & UPDATE
Retrieved Data
DATA REDUCTION & ANALYSIS
Output

retrieval apparatus, one should think of it as an integrated facility which is able to handle storage and retrieval as well as processing, manipulation, and finally, presentation of the data in one continuous process. It is for this reason that the segment at the bottom of the flowchart (see page 17) is provided for monitored analytical programs in which the retrieved data will become inputs, in turn, into analytical programs and their many routines and subroutines of which entire libraries have been accumulating. Here the newly retrieved data can be immediately subject to very efficient, even though complex, reduction and analyses. In addition to the more conventional statistical procedures in the direction of classical bivariate and multivariate analysis, time series decomposition can readily be undertaken as well as the more nearly experimental approaches along the lines of simulation, pattern search, and tentative excursions into the field of the generation of hypotheses. Put differently, cast in a system such as the one discussed here, manpower data could become a promising vehicle for truly imaginative, inventive, and daring explorations apart from their ready and convenient availability for the filling of more conventional analytical gaps.

In the casual practice of the occasional data-user, it would be quite conceivable to convey to such a system an information request by using simple, verbal statements to indicate what is to be done with the data once they have been retrieved. The alert and enterprising analyst, however, will find it rewarding to know more about the system and its operations. Learning the essentials of the art of communicating with a system would therefore be necessary lest the analyst depend completely and exclusively on interpreters and computer specialists for his conversation with the system. Since a system must be instructed concerning what to do with the retrieved information, the search statement should not only contain the request for the data to be retrieved but should also specify statistical or other operations to be performed on the data and, last but not least, the format in which the analyst wishes the information to be presented by the computer. Even so, it would not be necessary for the requesting researcher to be fluent in computer language because auxiliary programs in the form of appropriate compilers[3] can be furnished for translating his search statement into computer language. Within a particular system, this is the function of the so-called request-generator segment (see chart, page 17). However, it would be to the researcher's advantage to be familiar with the general nature and principles of computer operation, codes, and storage indexes so as to be able to exploit the system's capacities to the fullest. In other words, rather than being a fluent conversationalist, the data analyst should have a good grasp of the grammar and syntax of the system's language.[4]

To purchase information, so to say, rather than to windowshop for it in published tables and charts, the analyst must know precisely what it is he wishes to request. If this basic prerequisite were not so often disregarded, the requirement just stated might seem offensively trite and commonplace. Once, however, the prerequisite has been fulfilled and the analyst has decided on an articulate data request, the communication with the system is straightforward and in most instances considerably less laborious and time-consuming than most information search in the conventional sense of the term. Without concerning himself with the particular system's details, the prospective data-user can grasp the few language principles governing the dialogue between data-requester and information system quite easily; and dialogue it is, not figuratively but actually, because the properly phrased inquiry will be answered by the information system either in making the requested data available or in letting the inquirer know that the data are not available or that the desired data manipulations and computations could not be performed and, very often, the reasons therefor. Thus the researcher is greatly rewarded for acquiring a simple vocabulary and learning the syntax of the system's language.

---

[3] Also called "compiling routine," this is a set of prearranged instructions to make the computer translate a program in any arbitrary code language into machine code.

[4] The researcher should not confuse the system's language with programming language. The "conversations" here referred to would be between researcher and the "system"; they are in the system's language, and no further translation into programming language for use by the computer is necessary. A possible exception to this rule would be a situation in which the requesting researcher wishes the system to perform statistical or mathematical manipulations of the data which are not available in the form of subroutines. In this rather exceptional situation, where tailormade processing of the retrieved data becomes necessary, the analyst may prefer to be his own programmer.

V. Analytical Possibilities of the Information

The analyst of the information must be able to identify, locate, and secure access to the data; he must also be in a position to form an opinion about the quality of the data in order to arrive at a decision concerning their suitability for a particular analytical use. It is sometimes assumed that data qualification should also include problems of concepts and semantics. The analyst who derives information from CPS is nowadays in the rather fortunate position of being able to answer for himself most of the definitional and conceptual questions. The regular publications of CPS tabulations are and have been provided with ample notes to that effect, and other information of this nature is generally available.[1] Where the analyst faces definitional and conceptual problems which he cannot solve, he may use the detailed and extensive *Interviewer's Reference Manual* for the CPS and Housing Vacancy Survey field staff; the manual will help him form an opinion on the intent of the questions and on the way in which they were presumably asked. Unlike for some other statistical data, the conceptual content of the several questions can, as a rule, be quickly established. This does not mean that the particular terminology used is necessarily self-explanatory; as with any words in general, the analyst must first familiarize himself with the actual coverage and meaning of terms, captions, stub entries, and the like. It must be obvious that such a term as "unemployment," if not defined, would mean quite different things to different people; but there should be no difficulty for the analyst to find out the actual statistical meaning of the term "unemployment" as used in the publication of the results from the CPS. This does not imply that the respondent, in turn, is always fully cognizant of terminology and concepts or that his answers are always correct and valid. This, then, leads directly to the problem of qualifying statistical data not because of definitional or conceptual problems but because of errors inherent in the survey procedure — procedure in the broadest sense of the term including not only sampling but also nonsampling aspects.

The presence of errors requiring qualification of the data is particularly acute when the data are obtained through partial canvass, as in the CPS; but errors in the data are by no means limited to sample estimates. And so, although sample data are commonly suspected of sampling errors, the presence of nonsampling errors even when the data are based on complete enumeration may well be equally or more important. And since the results of an analysis can be no better than the raw materials which entered it, the quality of the statistical input is of decisive consequence. Ultimately, it is this quality of the data which codetermines the validity of the research results; and anything that can be learned about the information prior to embarking on its analysis can help the researcher to escape dangerous — although sometimes pleasant — illusions, waste of effort, and the temptation to draw spectacular even though invalid conclusions.

To guide the nonstatistical data-consumer through at least a part of the labyrinth of errors inherent in the quantitative information, the relevant features of CPS data are summarized in Appendix II, Qualification of the Data.

Data Supply and Demand

In the present sections, an attempt will be made to focus on the analytical use of manpower statistics. In this attempt the emphasis will be, as already mentioned, on that part of the supply of government-collected statistical information which has been furnished by means of CPS, and of the kind which is now usually published in the *Monthly Report on the Labor Force*. In other words, the just-described statistical information will be the information input, that is, the data supply which will be confronted with the demand for, and the analytical potential of, the information. Put differently, the problem here is to inquire into the interrelation between supply and demand concerning statistical data, and in particular into the degree to which available data are being analytically exploited.

As in any relationship in which supply is confronted with demand, the situations of greatest interest are the ones which reveal noticeable discrepancies because either the analytical demand for data exceeds their availability or a considerable amount of available information seems to find no appropriate demand and must therefore go analytically unused.

---

[1] See U.S. Bureau of the Census, ***Current Population Reports,*** Series P-23, No. 13, "Concepts and Methods Used in Household Statistics on Employment and Unemployment from the Current Population Survey," June 1964; U.S. Congress, Joint Economic Committee and U.S. Office of Statistical Standards, ***1964 Supplement to Economic Indicators: Historical and Descriptive Background,*** 88th Cong., 2d sess.; U.S. President's Committee To Appraise Employment and Unemployment Statistics, ***Measuring Employment and Unemployment,*** 1962, Chapter II; U.S. Congress, Joint Economic Committee, ***Unemployment: Terminology, Measurement, and Analysis,*** 87th Cong., 1st sess., 1961; and U.S. Bureau of Labor Statistics, ***Techniques of Preparing Major BLS Statistical Series,*** Bulletin 1168, 1955 (revision expected soon).

Primary interest in the data has, of course, been motivated by the role they have been playing as barometers and as measures of the manpower situation from month to month. Because CPS data reflect, like a mirror, the dynamics of the economy and its health status, they have a prominent position among the economic indicators prepared for the Joint Economic Committee by the staff of the Committee and the Office of Statistical Standards in the Bureau of the Budget;[2] their significance as a measure of the state of the economy when seen from the point of view of its dynamics is revealed by the distinct place they occupy among the National Bureau of Economic Research's types of "leading indicators."[3] In addition, there is inherent in the data a considerable analytical potential of a broader kind: the analysis of and research into the role played by manpower as a major national resource, its utilization patterns, its structural aspects, and the like; in other words, a problem catalog spanning most of the spectrum of the social sciences. Because of this breadth of topical implications of the data, because of their ~~remarkably~~ high level of quality and methodological refinement, and because they have been available for the population at large, on a monthly basis, for a longer period than nearly any comparable collection of statistical data, they are a statistical source without equal in this or other countries: they are, in this sense, truly unique. Therefore, their analytical potential extends far beyond their significance as an intelligence system for readily available indicators and "trigger" figures.

To form an opinion of the supply of CPS-generated statistical data, it first must be realized that the information which reaches the analyst through the conventional means of dissemination such as the *Monthly Report on the Labor Force,* the household data section of *Employment and Earnings,* and the detailed tables appendixes of the Special Labor Force Reports and of documents such as *The Manpower Report of the President* represents a relatively small part of the great and increasing mass of tabulations of CPS data which have been undertaken over the years. Especially over the past five years or so, that is, since the use by the Bureau of the Census of Univac 1105, the regular data output has been impressive.

In view of the generous size of these monthly tabulations undertaken for the Bureau of Labor Statistics, and the resulting hazard of obtaining "thin" cells, the reluctance to inform the public about the existence of the tabulations can perhaps be understood; efforts by a government office to forestall the danger of a flood of irrelevant and naive inquiries can be well appreciated. On the other hand, in the experience of many an institution of higher learning it has been found that the well-intended but often inadequate homemade survey as an information source is considerably less adequate — less adequate frequently from a survey-technical as well as from a sampling-error point of view. Also, some of the points made in defense of the restrictive policies with respect to dissemination of the information are less well taken if the use of the data is to be for research. Here it must be assumed that the analyst is going to proceed with due caution, and that even tentative and only suggestive findings can be of the greatest significance. They may generate surmises and conjectures which stimulate speculation and may turn into hypotheses and from there into propositions, models, and useful analytical constructs. Ultimately, it is the researcher who takes the risk, and it is the individual researcher's assumption of this responsibility that is the very foundation of any true research activity.

The tabulations now consist of "basic," "supplementary runs," and "gross change" tables. The "basic" tables supply information of interest to the analyst because, as their name suggests, they underlie practically all the published table material. Only a relatively small part of the basic tables material finds its way into publication, and that which does as a rule is in much less detail, usually because of the size of the sample.

In the same sense, the so-called "supplementary runs" provide general data, although here even less of the available information finds its way into published and generally available documents.

The third group of regular monthly CPS tabulations are the 12 "gross change" tables. Gross changes, sometimes called "gross flows," refer here to intermonthly changes in employment status of one and the same individual.[4] In this sense gross change data result from micro pattern as against "net changes" which are a macro phenomenon — i.e., the net balance of inflow into (or outflow from) an entire labor-force-status category after having accounted for offsetting movements in the opposite direction. The great interest in gross flows data results from their being a measure of the dynamics of the labor-force-status patterns,[5] and from their being measures on the lowest level of aggregation, i.e., that of the individual himself. This aspect of gross flows data is also a cause of their limitation.[6] Since the data are derived from CPS, a sample survey with fixed primary sampling areas (fixed within the restriction imposed by the rotation feature[7] of the design), the chain of pairwise monthly comparisons is interrupted for movers out from such a sampling unit — or, for that matter, movers into a sampling area. Furthermore, interview failure because of noninterview has here a cumulative effect since its effect is proportional to the sum of noninterviews for nonidentical persons, and of noninterviews of identical persons in two successive monthly surveys. The problem is an irksome one but not one that destroys the analytical potential of the data.[8] The Census Bureau has been keeping the problem under observation; and in view of the importance of gaining better insight into the dynamics of the labor force, the entire question of improving data and refining methods to study these gross flows deserves attention as an object for analytical utilization of CPS data.

---

[2] See U.S. Congress, Joint Economic Committee, *Economic Indicators,* various issues; also *1964 Supplement to Economic Indicators, op. cit.*

[3] See, for instance, Geoffrey H. Moore, editor, *Business Cycle Indicators,* Vol. I, *Contributions to the Analysis of Current Business Conditions,* and Vol. II, *Basic Data on Cyclical Indicators,* studies by the National Bureau of Economic Research, Princeton, Princeton University Press, 1961.

[4] See U.S. Bureau of the Census, *Current Population Reports,* Series P-50, "Gross Changes in the Labor Force: May 1948 to January 1949," and Series P-50 annual reports 1949-1952; a succinct summary was given more recently in U.S. Bureau of Labor Statistics, "Gross Changes in the Labor Force: A Problem in Statistical Measurement," by Robert B. Pearl, *Employment and Earnings,* April 1963.

[5] See pp. 31 and 33, this study.

Here it is merely pointed out that there is a wealth of suggestive data accruing every month that bears considerable analytical promise. Up until 1953, the Census Bureau published summary data regularly;[9] after that date, publication was stopped because of a number of problems such as the ones mentioned above; however, the tabulations of data pertaining to gross flows have continued. It should perhaps be pointed out also that the President's Committee To Appraise Employment and Unemployment Statistics recommended "that the publication of gross-change data be resumed" and stated its belief that "despite the deficiencies these tabulations could be useful for research."[10]

Concerning CPS information supply, in retrospect one finds that:

1. There has been tabulated and cross-tabulated material far in excess of what is published.
2. If an information-storage system were developed, it would enable the analyst to obtain the best and most practical data inventory.
3. In the meantime, the Univac tabulation specifications convey a general impression of the data supply. However, they do not give an adequate picture of the vast store of information derived from arranging the data in sequence of time; a description to that effect must await the completion of a data inventory such as that mentioned above.

And not one of the so-far-mentioned points has even touched upon the wealth of micro information, which yields a set of completely flexible building stones and components which "in principle," could be arranged in any desired form and combination.

What, then, are some of the analytical possibilities inherent in this mass of data? It should be no surprise that, given this rich and ample diet, by no means all of it has found appreciative takers; data gourmets seem to be rare in this particular area, and they have a constant fear of indigestion. Naturally, there are also situations where appetites for special data must go unsatisfied for lack of desired information.

Looking at the data store as the supply side, and at the analyst's prospective data requirements as the demand side, the three states that might be construed as typifying the demand-supply situation are: Demand may equal supply; it may fall short of it; or it may exceed it. The latter two situations are here of particular interest since they reflect imbalances which need study with a view to remedial action.

It is tempting to think of the implied two-way relationship between supply of and demand for data in terms of a two-way table where data outputs form one dimension, and data inputs required by analysis-oriented activities form the other. However, there seems to be no hope at present to fill in such a table, even with synthetic quantities.

On the analysis-oriented data requirement side, a first step would have to be the development of an operational taxonomy, useful for the study of the relationship between statistical data supply and demand. In the absence of an indexing system, a grossly oversimplified grouping of analytically oriented activities requiring data inputs will be used as a general framework within which to sketch the contours of the data demand and supply picture.

In considering the requirements for specific kinds of analysis of CPS manpower data, one must remember that the restriction to CPS data is here a self-imposed one, and that in actual research practice there often is more than one set of data for use in a given analytical problem. On the other hand, it may be well also to recall that CPS data are the only broadly based data which pertain to the entire population, within and outside the labor force, and to all components of the labor force without regard to selective occupational or industry groupings and the like — that is, without regard to administrative and legal restrictions of coverage which often apply to other sources of manpower information. Compared with the population censuses (the only other source of information on the entire population), CPS data have the advantage of being a continuing and current source of information. On the other hand, they are less complete as to content; they are subject to sampling errors and are generally beset with smaller nonsampling errors. All this goes to show that the ultimate decision concerning what data to use for which research objective, and in what way, must rest with the individual analyst. Here consideration is given to all such analytical potentials of CPS data that seem to be promising if CPS data are, or will be, available in appropriate form for use by the analyst.

---

[6] See Appendix II, p. 57, this study.

[7] See Chapter III, p. 8, this study.

[8] For instance, geographic mobility within the sampled households is approximately 1.5 percent per month; the noninterview rate is between 4 and 5 percent per month with a peak of 6 percent in summer when the not-at-homes contribute their largest number of noninterviews. See Pearl, *Employment and Earnings, op. cit.*, p. vii, and sources there quoted.

[9] U.S. Bureau of the Census, *Current Population Reports*, Series P-50 and P-59.

[10] U.S. President's Committee, *Measuring Employment and Unemployment, op. cit.*, p. 16.

The analytical uses referred to will be grouped into the two major sectors: substantive analysis, and methodological utilization of the data. Substantive analysis must account for the more nearly descriptive uses of the data and also the generalizing, theoretical, "modelizing" approaches. With the arrival of the electronic computer, the once-established distinction of what used to be referred to as "idiographic" (for the descriptive), and "nomothetic"[11] (for the more nearly analytical function of statistics) is becoming less incisive; in the actual practice of the data analyst the two approaches or goals have sometimes become operationally indistinguishable. Deplorable as this tendency to merge the two approaches may perhaps appear from any point of view but that of extreme empiricism, it is here accepted as a fact in analytical practice; and no attempt will be undertaken to draw fine lines between the two.

With respect to the second major category of demand for information, i.e., information to be used for methodological purposes, it is believed that a great deal may be accomplished along the lines of studying and improving method, measurement, and conceptual framework; and also that such analytical uses of the data must be included along with the more nearly content-oriented ones. Nearly all the recent committee findings, recommendations, and criticisms point in that direction.

Suitability of the Data for Substantive Analysis

Returning now to the use of CPS data for substantive analysis, it appears that for purposes of presentation a convenient way of grouping various analytical efforts would be according to the two principal categories of data inputs into micro and macro data employing analysis. In a similar way, one could order these analytical efforts also according to output or major type of analytical activity into cross section and longitudinal studies.

The division into micro and macro data, formalistic as it may seem, is of some practical consequence for the analyst. In principle, any macro statistics can be obtained from the corresponding micro information. However, because of the cost of tailormade tabulation and estimation, because of various administrative hurdles, and because of the problem arising from disclosure policies and practices,[12] the analyst will as a rule have recourse to micro data only where appropriate macro data cannot be found or where they are not adequate. For instance, he will resort to micro information where:

1. Matching of CPS data for individual persons or households is to be undertaken with other kinds of source documents, e.g., Internal Revenue Service or Social Security Administration records.
2. Tracing through time is involved of records for individual persons and households.
3. Direct statistical manipulations are envisaged such that aggregating and grouping original data would result in unwarranted loss of information.
4. Indirect statistical manipulation of the source material is required because desired macro data are not directly available and must be composed afresh from basic micro materials.

Another way of grouping various types of CPS data demand would be according to the content emphasis of the intended analysis or description. It was found that among the more articulate contemporary streams of thought, styles, and preoccupations there could be recognized the following, not necessarily mutually exclusive contents:

1. A tendency toward disaggregation to which there corresponds in the field of economics a growing recognition of micro economics.
2. The penchant toward the study of flow and change patterns, of trends — in other words, the study of dynamic problems.
3. The emergence of new taxonomic perspectives, questioning some of the traditional concepts and classifications.
4. The growing interest in global models and descriptive schemes and systems.

This study follows, more or less, this latter subject-matter-oriented approach and at the same time tries to superimpose some of the former classification relating to the micro-macro classification, and the cross-section-longitudinal dichotomy.

It is, of course, not practical even to attempt a complete presentation of analytically oriented demand for CPS data. The cases listed on the following pages are therefore intended to be illustrative rather than exhaustive; they are based on conversations and interviews, on correspondence, and on direct or implied references in current literature; they are noted here because they create a demand for information and statistical data that could be satisfied, in part at least, by using CPS results.

---

[11] At the time the customary terminology, created by W. Windelband and introduced into statistical literature by A. A. Chuprow, "nomothetical," also "nomological," applied originally to the establishing of "laws" and of "causal relationships," whether deterministic or probabilistic.

[12] See how such hurdles may be overcome with respect to Census data in a paper by Morris H. Hansen, *Some New Developments in Increasing the Availability of Census Data,* presented at the meeting of the American Statistical Association and American Economic Association in Chicago, December 30, 1964, 9 pp.

Analysis on Low Levels of Aggregation

Concerning the utilization of CPS-type information for analysis on lowest levels of aggregation, the following should perhaps be mentioned first: There can be observed a certain degree of parallelism between micro economics and cross section analysis on one side and, on the other, between macro economics and time series or longitudinal analysis. This parallelism, such as it is, has been a historical phenomenon, and has been due to the prevailing characteristics of the available data input rather than to logical cogency.[13] Ideally the two approaches are not exclusive but reinforce each other; however, statistically efficient methods for pooling the two are not often available.[14] It would seem that the pendulum is so far out toward the macro perspective, especially in the general area of manpower analysis, that a fuller exploration of micro approaches would appear well justified and promising. Although logically not identical, but of a nature similar to micro information, are data on low levels of aggregation; therefore, they will also be included here.

The general tendency toward disaggregation down to the micro level can be readily understood. With growing analytical refinement, and with increasing intricacy of problem situations in labor market and manpower analysis, emphasis in empirical and theoretical studies has been gradually shifting toward disaggregation. For example, interest has been veering toward individual behavior, toward the household, the particular industry, occupation, labor-force-status category, geographic area, and away from broad aggregates such as total employment and unemployment. Although this tendency can be observed more clearly in areas other than manpower and labor,[15] it is nevertheless clearly present in these fields; and it is likely to increase in usefulness. Last but not least, the normative parts of micro economics, known under the name of welfare economics, are coming again to the fore as the social aspects of the unemployment problem and the economic aspects of the manpower utilization policy are commanding increasing attention.

Among the many problems whose solution might benefit from disaggregation of data has been the complex of questions involving educational level and its relationship to unemployment, to technical change, to the adaptability of the labor force, and to varying labor force participation rates.

Although the CPS control card, which is filled in when a household is first included in a given sample, does have data on educational attainment,[16] the conventional source for this kind of information is the March supplement to CPS. The last survey of this sort was undertaken in March 1965; previously surveys were made in 1964, 1962, 1959, and 1957. Answers to the supplemental questions are now tabulated and published in the corresponding Special Labor Force Report under the title *Educational Attainment of Workers*.[17] Most of the tabulated materials were actually published in the appendix to the Special Labor Force Report No. 30, but there exists information in somewhat greater detail, for instance, with respect to age distribution.

In addition, it is of course not at all inconceivable to fall back on micro data where already tabulated information is on too high a level of aggregation. Among the great number of tempting analyses are those relating educational level to job mobility while holding constant the influence of industry, occupation, age, sex, and color. In other words, here as elsewhere, it would not seem irrelevant or inappropriate to experiment with the use of tools taken from the toolkit labeled, "multivariate analysis," in order to gain further insight into the complex interrelations between educational level, employment status, and a host of socioeconomic factors.

It should perhaps be stated at this point that what at first glance might seem an unusually high cost of processing data would very often be found to be quite reasonable if contrasted with the cost of a comparable tailormade and *ad hoc* survey. This applies also to most analytical problems other than those connected with the educational attainment question. The issue here is not really the "processing cost" but whether it is desired to undertake the particular research and what its value is to the scientific community or to the community at large.

Among analytical efforts pertaining to disaggregation of broad national totals into particular manpower categories, perhaps the most persistent have been those referring to women. In spite of this, there remain a great many open questions and challenging analytical problems. Even additional comparisons of the labor force behavior of married women with that of single women might reveal further interesting patterns — so would inquiries into the role of women as contributors to family income, and the changes over time of their role as contributors to family income.

---

[13] See, for instance, Lawrence R. Klein, editor, *Contributions of Survey Methods to Economics,* New York, Columbia University Press, 1954, p. 194. "As a matter of practical application, cross-section analysis of survey data has usually dealt with micro-economic behavior, while time-series analysis has usually dealt with macro-economic behavior. This difference is not a necessary consequence of the two approaches but is likely to occur frequently as long as the collecting of data continues in its present patterns."

[14] On the problem of pooling the two approaches in general, and on disaggregation in particular, see, for instance, Lawrence R. Klein, *Econometrics,* Evanston: Row, Peterson & Co., 1953, pp. 185 ff., 221 ff., 265 ff.; Stefan Valavanis, *Econometrics,* New York, McGraw-Hill, 1959, p. 5; Gerhard Tintner, *Econometrics,* New York, Wiley, 1952, p. 109, and literature there quoted.

[15] J. M. Henderson and R. E. Quandt, *Micro Economic Theory,* New York, McGraw-Hill, 1958, Sections 1-2.

[16] For persons 25 years and over, questions 21 and 25 attempt to establish highest grade of school ever attended, and year that the particular grade was finished.

[17] Special Labor Force Reports, *Educational Attainment of Workers,* No. 53, March 1964; No. 30, March 1962; and No. 1, 1959; earlier surveys were published as *Current Population Reports,* Series P-50, Nos. 14, 49, and 89. Results of 1965 have not yet been published.

Another example is the analysis of labor force participation by women according to the number of young children in the family, of the existing differentials, and the changes over time of these patterns as shown by their cyclical response and their longer range component. Here the March supplements on family characteristics yield interesting information on the labor force status of wives, by income of husbands; on the labor force status of women, by age and presence of children by broad age groups; and, if simultaneous information on income is not needed, by rather fine age groups. Much, but not all, of this information can be reconstructed from the appendix tables to Special Labor Force Reports Nos. 2, 7, 13, 20, 26, and 40. For statistically sophisticated analysis, especially where multivariate relations are to be explored, micro data might be needed; but a good deal of the now available macro data are on sufficiently low levels of aggregation to permit interesting experimentation. The same sources also contain much material for a fuller exploration than now available of the inverse relationship between husband's income and wife's employment rate, although there may be involved some record-matching or retabulation of micro data where specific interrelationships are to be disentangled which involve age, occupation, and family size.

As a final illustration, the question might be explored whether or not and, if so, to what extent higher unemployment rates among women could be explained by differences in employment mix of men and women, by differences within the same occupation, and by differences within the same industry. Questions of this nature could be broached by exploiting the basic CPS tabulations on persons at work by full-time or part-time status, reasons for part-time work, hours of work, occupational group, and sex; and on unemployed persons by duration of unemployment, industry group, class of worker, color, and sex. For unemployed *and* employed persons tabulations might be exploited by industry group, class of worker, marital status, and sex. Again, for more subtle multivariate analysis, it may be necessary to have recourse to the construction of data on low levels of aggregation; this may become feasible through retrieval from the earlier described BLS information-storage system or, if necessary, from special-purpose cross-tabulation of information from microtapes.

As to disaggregation of broad labor force and population totals into finer subcategories, the three which have attracted perhaps most attention are the nonwhites, the old, and the young. Much material pertinent to the three groups has been collected and tabulated, and a good deal of it has been made generally available. However, most of this information has been used predominantly for descriptive purposes; only rarely have the data been subject to more complex, hence frequently risky, analysis.

For instance, in view of the wealth of data describing individuals belonging to these three subcategories, it would seem quite feasible to venture at least experimentally into the field of discriminant analysis;[18] in addition to ascertaining the nature of the discriminating planes, the intensity of the separation could be studied in terms of the residual probabilities of misclassification, in order to form an opinion as to what circumstances and factors are the most likely to contribute to the separation.

In addition to this kind of analysis, which is predominantly cross-section-oriented, there is genuine interest also in longitudinal analysis on the micro as well as the macro level. Among the problems falling into this group are those concerning the dynamics of the labor market and the labor force. The use of micro data would be of particular promise if such questions were to be explored as "voluntary" versus "forced" withdrawal from the labor force. Here it might be rewarding to consider the feasibility — perhaps on the small scale of one or a few selected CPS primary sampling units — of implementing the basic CPS information with interview specifically aimed at the withdrawal motive.

A high priority is being held by the urge toward geographic disaggregation — disaggregation into local areas or into geographic strata. But there seems to be a good deal of misapprehension concerning data requirements for such analysis.

---

[18] Discriminant analysis (discriminatory analysis in British usage) is in essence a statistical method not unrelated to regression analysis, applied to the taxonomic problem of classifying an individual observation as belonging to a particular population or predetermined taxonomic class while minimizing the probability of misclassification. The analysis can be particularly illuminating when only two classes are involved, and when the problem therefore is to separate two groups for whose individual members measurements of several variables are available. Although the computations are tedious, especially if a large number of variables are involved, programs for high-speed computers make the problem into a routine operation (see, for instance, William W. Cooley and Paul R. Lohnes, *Multivariate Procedures for the Behavioral Sciences,* New York, Wiley, 1962, Chapter 6, and bibliographic references on p. 133). When differences between pairs of groups are to be analyzed (for instance, employed and unemployed, male and female members of the labor force, old and young unemployed, etc.), discriminant analysis may yield suggestive results if used directly (for the classification of individuals) or indirectly (e.g., to find out whether a given set of suspected characteristics actually produces a significant separation, and which of the several variables is likely to account for most of the discrimination between the two groups). See, for insance, M. G. Kendall, *A Course in Multivariate Analysis,* New York, Hafner, 1961, pp. 144 ff.; for an early application to economics, see David Durand, *Risk Elements in Consumer Installment Financing* (Technical Edition), National Bureau of Economic Research, New York, 1941, especially the appendix for a not-too-technical discussion of the method.

Because CPS is a sample survey, the attrition of the evidence following geographic, as any other, disaggregation will be felt quite markedly especially when analysis is to be undertaken of several variables jointly and within small geographic subareas. In other words, case studies of such geographic subdivisions are now, in general, not within the scope of CPS.

On the other hand, a great deal of analytical potential of the data might be realized toward geographic disaggregation if the analysis could be aimed at entire strata (not necessarily contiguous) of locations with, say, common socioeconomic characteristics. If such groupings by place of residence are relevant at all to the exploration of a particular labor force problem, the collection of a number of sampling areas, although individually small, may result in agglomerations large enough to make analysis well worth while. Thus, the researcher's task is to identify promising problems which would lend themselves to the study through geographic groupings now within the statistical scope of CPS. The suitability of the data for such analysis might be enhanced by aggregating the data over time, e.g., by considering annual averages based on 12 monthly observations rather than single months. Thus, it would seem that there is here sufficient ground for experimentation by an enterprising analyst. Also, it should not be forgotten that increases in the sample size of CPS have been and are being seriously considered.

The above suggestions in no way exhaust the interest in disaggregation of overall information, or the interest in the micro aspects of the manpower problem; rather, they serve to point to a possible use of the sheer limitless store of micro data, and of the more readily accessible data on low levels of aggregation in terms of the smallest "building blocks" resulting from any of the basic CPS tabulations and cross-tabulations. The resulting analytical potential becomes even more impressive if one realizes that these building blocks have been produced, month by month, or year by year, as far back as the particular CPS information has been collected and tabulated.

Analysis of Special Taxonomic Groups

Whereas the preceding cases referred to analytical problems requiring information resulting from disaggregation of large global composites, the situation now referred to exemplifies to some extent the opposite analytical interest — the attempt to learn more about manpower and related problems by considering an entire taxonomic group of particular analytical interest, e.g., the family or household rather than its individual members. A great many socioeconomic problems and behavior patterns can be understood much better if viewed in terms of the family as a unit of observation. The primary source for this information within the CPS framework is the March Supplemental Survey on Family Characteristics and on Marital Status of Workers. The corresponding data were published for 1960, 1961, 1962, and 1963 in the Special Labor Force Reports, Nos. 2, 7, 13, 20, 26, and 40. Much more information, and often on a monthly basis, can be obtained on various labor force characteristics by marital status. However, the only inference that can here be drawn with respect to family is that based on statistics referring to husband-wife relationships "with both present in the same household." The March supplement on family characteristics provides tabulation by labor force status, by husband's and by family income, by contribution of wife to family income, by broad age-groupings of children, by age and occupation of wife and of husband, by composition of family in terms of broad labor-force-status groups of family members, and by employment status of head of household.

Again, the present use of the information is primarily descriptive, with few analytical overtones. Most exploited for analytical purposes are the two tabulations which illuminate the interrelationship of family income in a given year, wife's work experience during the preceding year, and wife's present earnings. Micro data, if used directly for analysis, would make it possible to bring to bear more forceful statistical methods which could utilize the empirical information more efficiently. For instance, even mere regression of family income on household head's income by various subgroups standardized, say, for family size and composition, for occupation, and for age of spouse, and the like, might disclose the existence of patterns whose relative invariance over time could be ascertained from the successive annual surveys.

After some additional tabulations and retabulations of the survey results, at least a tentative analysis could be undertaken of labor force participation in its relation to family characteristics now available on the schedule, in order to learn more about which characteristics are likely to be more, and which less, important determinants. This could be attempted, if two- or three-way presentation of tables were not a requirement, by considering the several characteristics and variables simultaneously, in order to form an opinion as to their independent and joint relevance, if any. Because of the actual or suspected existence of strong correlations — correlations which are deeply embedded in our social and economic fabric — of, say, family income with labor force status of wife and other family members, age of family, color, and geographic factors such as the urban-rural break, a two- or even three-way tabulation is unlikely to show anything but the most obvious relationships; also, because such a procedure has a tendency to obscure or veil second- and higher-order relationships, it may encourage the acceptance of spurious associations. It is this danger that the analyst must weigh against the convenience, straightforwardness, and ease of communication of only two- or three-way cross-tabulations.

On the other hand, the relatively large number of available variables and attributes suggest that an investigation along the lines of component and factor analysis might[19] be considered in order to ascertain whether all the variables are really needed and, if not, which variables the analyst could "get along with" without endangering the validity of the end results.

In essence, the just-discussed problem of using the family rather than the individual as a unit of observation in the explanation of certain manpower problems is a taxonomic one. There are other similarly taxonomic questions which can be approached by using available CPS materials; there is, for instance, the problem of the part-time as against the full-time worker. Unlike the illustration used in the preceding section where the primary source of CPS information was the answers to supplemental CPS surveys, once a year or even less frequently, here the data are derived from the basic CPS schedule itself; therefore, they become available once a month.

Several analytical opportunities come to mind once emphasis is placed on "intensity" of work or on "propensity to work"— in terms of time devoted to the job —as a taxonomic criterion. The pattern of part-time work and its changes over time are of some interest because present publication practice[20] limits the information to aggregative figures in terms of the estimate of all employed persons who, during the survey week, worked part time for noneconomic reasons (i.e., because of bad weather, industrial dispute, vacation, illness, holidays, and other such reasons) and economic reasons. A separate estimate is given for those part-time workers who usually work full time. Since "part time" is defined as "less than 35 hours per week," and since the number of hours worked refers to "all jobs," the distribution and relative stability or variability of hours worked are of interest. Also, the differences in these patterns between various exposure groups, e.g., between males and females, and among selected nonagricultural industries would seem to be of further analytical interest and promise. Another question of considerable importance for the interpretation of the economic and social significance of part-time work is the one which relates part-time work to marital status and age of part-time worker, by major category of "reason" for part-time work. An analogous inquiry into the reasons for part-time work among those who indicated that they usually work full time compared with those who usually work part time would greatly help in identifying the regular part-time workers, and in separating them from the occasional ones. Although these and similar problems have, of course, attracted the attention of labor economists and others for some time, it would seem that further exploratory analytical efforts might be rewarding if data were available in a format sufficiently flexible and adaptable to fuller exploitation.

So far, the suggested analyses can be undertaken with macro information, "in principle" available now. To gain fuller insight into the phenomenon of part-time work, and possibly into some of the factors determining the number of hours worked in general, it is desirable to resort to micro data; this seems to be indicated in at least two situations: (1) the preparation of special-purpose cross-tabulations; and (2) the additional interviewing of subsamples of households from selected primary sampling areas, in order to learn more about "reasons for" and "circumstances surrounding" the particular part-time work. The special-purpose tabulations of micro information would be aimed at reconstructing the relationship of part-time work with household composition, family income characteristics, and selected demographic characteristics so as to round out the comparison of profiles of part-time workers' households with those of full-time workers — thus supplementing the presentation of the above-mentioned macro information, and attempting to separate the more nearly "involuntary" from the "voluntary" part-time workers.

Another problem, also taxonomic in nature, is the definition of part-time work. In present statistical practice, "part time" stands for working 34 hours or less per week. It has been pointed out that this need not be the only way of describing part-time work. To the extent that "normal" — that is, perhaps some "average" — weekly hours of work could be established for subgroups of the employed labor force, homogeneous for the particular purpose, part-time work might be defined in terms of negative deviations about the subgroup's specific norm. Beyond this, a question also arises whether the exposure period for which the hours-worked patterns are to be analyzed should be a week, as in present practice, or some other period of time — perhaps a month, a year, or theoretically an entire working life. In terms of unemployment, there are two meanings associated with part-time work: a less-than-full-income-earning capacity, and a less-than-complete utilization of the manpower resource; because they raise two different questions having quite different implications, one should expect different answers to them. This would seem to suggest the possibility that different definitions or measures of part-time work may have to be applied in the analysis of the respective problems. Accordingly, taxonomic plurality may here be more advisable than strict uniformity of treatment. Since "hours worked" are tabulated by much smaller intervals than less-than-35 and 35-or-more hours, there is considerable scope to experiment with various part-time concepts. Valuable insight may further be gained by inspection of information contained in the February supplements data on so-called work experience, beginning with 1947. This supplemental information refers to the number of weeks that the interviewee worked, full time or part time, during the given calendar year; to the number of weeks during which he was not working, not looking for work, and the like. Thus, aggregative data are available, and many of them are published in the statistical appendixes to the Special Labor Force Reports quoted in footnote 17 of this chapter on part-year workers in addition to what, in fact, are part-week and part-day workers.

---

[19] Component analysis and its converse, factor analysis, are here referred to as representative of the large class of multivariate techniques and methods. Such multivariate procedures are of particular interest within the present context for two reasons at least: First, CPS can be a major source of multivariate information if the data are made available in sufficiently flexible form and freed from the restrictions imposed by conventional tabular representation; and second, the data-storage system and its close connection with electronic data-processing equipment raise the hope that an accessible and practicable path will be found to the kind of numerical analysis which is involved in most multivariate techniques, and which frequently takes the form of data manipulations along the lines of matrix algebra for which high-speed computers can be very efficiently used. Data-processing techniques which are involved in some of the more complex but powerful analytical procedures, full of promise for use in the social sciences, would thus seem to be within reasonable reach of the analyst. See, for instance, William W. Cooley, and Paul R. Lohnes, *Multivariate Procedures for the Behavioral Sciences,* New York, Wiley, 1962, which conveys a good view of the relation between such procedures and electronic data processing; or M. G. Kendall, *A Course in Multivariate Analysis,* New York, Hafner, 1961, with numerous bibliographical references. For a more advanced treatment, see T. W. Anderson, *Multivariate Statistical Analysis,* New York, Wiley, 1958.

[20] See primarily Table A-14, various issues of U.S. Bureau of Labor Statistics, *Employment and Earnings.*

Again, a mere description by color, sex, age, household composition, occupation, and industry might greatly assist in gaining an insight into predominant work habits, with respect to time, of the American labor force and its components; such a description may suggest the presence of less uniformity than perhaps expected, and would go a long way toward establishing analytical norms against which specific profiles could be projected in order to classify them into "less than" or "more than" categories. The resulting taxonomy, if applied to economic and demographic characteristics of the persons involved, would also be helpful in shedding light on the earlier raised question of the social and economic implications and meaning of different time patterns of work.

As a last example for the analytical use of part-time-work information, the longitudinal analysis of macro-type information on time patterns of work should be mentioned. To the extent that part-time work as defined describes a state of deviation from normal, it becomes tempting to investigate the cyclical and seasonal deviations of part-time from full-time work; and to the extent that a sufficiently long series of data can be accumulated, there is also an opportunity to study structural change. The problem can be conceived of a classical time series problem in what are essentially autocorrelative terms of reference, or as a propagation problem in what is substantially a serial correlation approach. Best, of course, is the benefit from both kinds of approach. The nature of much of the information, the fact that the data are mostly monthly, and the fact that many of them can "in principle" be readily retrieved would make even the somewhat more complicated and elaborate techniques of time series analysis readily usable. For instance, if it were expected that within specific homogeneous subgroups of the labor force there existed a pattern relating unemployment to full-time work and, in turn, to part-time work where part-time work might be a particularly sensitive series, it would not be too difficult to establish analytically the fact of such a pattern, its nature, and the probabilities of particular lead-lag sequences among the three levels of labor force status: full-time, part-time, and no work.

By analogy, the lead-lag patterns could, in turn, be analytically arranged into sequences; for some labor force subgroups these patterns would turn out to be more sensitive to overall change than for others. Thus, the taxonomic problem that may have had its origin in the researcher's urge to understand better the labor market may have significance beyond the confines of the immediate problem in anticipating changes in the economy as a whole.

As a last illustration, a disproportionality may be pointed out in the statistical and analytical implementation of the basic classification of the population as "in" the labor force or as "outside" it. There is much more information available on those in the labor force than on those not in the labor force; the reasons therefor can readily be found in recent social and economic history and in the dominant role of the unemployment problem. From the point of view of manpower utilization, however, those outside the labor force, the "unutilized" ones, represent in essence as much of a problem as those within the labor force, and much more should be known about them. In addition, the taxonomy is here not sharp — at least in practice. The labor force and the non-labor-force sectors of the population are not always easily defined, and lines of demarcation are frequently fluid. For instance, when does an unemployed person, hence a member of the labor force, cease to be a part of the labor force? How exactly should an intermittent worker who does not seek continuous work be classified during various stages of his work cycle? Is the separation into labor force and non-labor-force analytically efficient in all important problem situations, or should the analyst think in terms of an underlying continuum extending from absolute reluctance to join the labor force to a maximum propensity to do so?

Where questions of the latter kind are involved in the analysis, much too little information is now available to estimate supplemental manpower reservoirs or to form an opinion on where along the just-mentioned continuum a particular person or group of persons stand so as to determine the kind of inducement necessary to make them move voluntarily into the labor force. However, it is suspected that even the small amount of data is not fully exploited.

Of particular interest here are micro data or such tabulations that, in fact, represent micro behavior as do the gross change tables. These data, to the extent that they mirror the movements into and out of the non-labor-force group, permit at least some analytical conjectures. For the most general macro analysis, especially analysis of a time series type, changing labor force participation rates may be of interest where age group differentials are at stake; and exploration of suspected patterns according to major types and classes of location may well be worth while. Here also the changing interrelations over time between unemployed and those outside the labor force can be suggestive — to the extent that they may be considered indicative of the forces "attracting into" and "driving out from" the labor force. For any more searching analysis, micro information will be preferable. Even the gross change tabulations, which are limited to relatively short periods of time, often affect the "not in the labor force" component more seriously than the rest. This will be of significance to the analyst who wishes to take a longitudinal point of view. Tracing of simple flow patterns and ferreting out even two- or three-dimensional shifts should be rewarding in view of the paucity of explorations now at hand. On the other end, more complex stochastic-process[21] models could be considered, conceiving of the passage of persons from one "state" (for instance, the non-labor-force subcategory, a specific industry, the employment status) to another as transition probabilities in a succession of matrices. Should it turn out in a particular instance that the transitions are more frequent and pronounced than expected, and that they display a certain stability over time, simulation models might be constructed to trace expected changes in the labor force and non-labor-force sectors of the population, according to specific stimuli introduced into the model for exploratory and experimental purposes. Conversely, such a model would assist in the speculation as to what kind of stimulus or inducement would be required to produce a given redistribution.

Modern electronic data processing is bringing with it the possibility of still another application of analysis to taxonomy which should be mentioned. Apart from tabulating data, and analyzing distributions and relationships according to various analytically promising taxonomic groupings, it is possible also to consider the reverse approach — the determination of efficient and analytically suitable taxonomies on hand of the observational data. In other words, given empirical information, the problem may be thought of as one of searching for proper, economically meaningful classification; thus taxonomy is reached by induction and posteriorically. Very little has been done along these lines in the past, and the nature of the problem would seem to point in the direction of approaches and methods related to pattern search.[22] It is therefore an area particularly suited to electronic data processing, and it would seem appropriate for the handling of information such as that derived from CPS. There lies here an entire region of analytical prospect, which is partly unexplored, but which is in no way limited to or uniquely relevant to CPS. However, because of the possibility that a statistical data-storage and -retrieval system might be developed, and because of the volume and nature of the data, CPS information would seem to be a particularly noteworthy vehicle for such reconnaissance and scrutiny.

---

[21] Broadly speaking, the theory of stochastic processes corresponds to the dynamic part of statistical method. Although the term "stochastic process" is synonymous with "random process" in general (see, e.g., William Feller, *An Introduction to Probability Theory and Its Applications,* New York, Wiley, 1950), its everyday use emphasizes especially the time dimension in which the process is taking place. An important category of such processes is the so-called Markov chains. They can be manipulated much more easily than the "general" process, and they seem suitable for modelizing observed time sequences in the behavioral sciences (see, for instance, Isadore Blumen, Marvin Kogan, and Philip J. McCarthy, *The Industrial Mobility of Labor as a Probability Process,* Ithaca, Cornell University Press, 1955; or F. Harary and B. Lipstein, "The Dynamics of Brand Loyalty: A Markovian Approach," *Journal of the Operations Research Society of America,* Vol. 10, January 1962). Such processes can serve as promising model components, in particular where it is not unreasonable to assume that "the future development is completely determined by the past state and is independent of the way in which the present state has developed..." This, however, does not imply that *unique* determination is necessary since availability of probability distributions is sufficient to make predictions (Feller, *op. cit.,* pp. 337 ff.) or to compare an observed process with an expected or otherwise postulated one.

For a simple presentation of a special case of particular interest to the social scientist, see J. G. Kemeny and J. L. Snell, *Finite Markov Chains,* New York, Van Nostrand, 1959; see also the bibliographical references in E. Parzen, *Stochastic Processes,* San Francisco, Holden-Day, 1962, pp. 307 ff. For an early treatment by an economist, see H. Wold, *A Study in the Analysis of Stationary Time Series,* Uppsala: Almquist & Wiksells, 1938. For an application to manpower problems, see Blumen, Kogan, and McCarthy, *op. cit.*

[22] The problem of pattern recognition is closely associated with the problem of taxonomy (see the not-too-technical treatment by M. Minsky, *in* Edward A. Feigenbaum and Julian Feldman, editors, *Computers and Thought,* New York, McGraw-Hill, 1963, pp. 406 ff., especially p. 411). In essence, so-called pattern recognition and search theory can be looked at as a further step toward reduction in the process of reduction of data initiated by statistics; in pattern recognition and

(Continued on page 29)

Dynamic Analysis

A third type of analytical interest can be assembled around the topics of economic and social dynamics of the labor force and of manpower in general. They can be conveniently subdivided into problems of "dynamics in the large" and of "dynamics in the small."

From the very inception of CPS, "dynamics in the large" has been a central theme underlying justification and reason for the very existence of the entire CPS program. Early attempts to identify the emergence and seriousness of an unemployment problem also fall primarily into this group of analytical interests; so do later attempts to establish warning systems which would gauge imbalances in the labor market.

And because there has been, and still is, a great deal of analytical interest in economic time series, this fact alone is exerting strong pressure and demand for the information; the particular research is likely to range all the way from the more conventional forms of time series analysis to, say, spectral analysis which is of potential practical interest also in connection with the elimination of the seasonal component from CPS data.[23]

In view of the huge number of available CPS series, it is perhaps surprising that few of the data have been subject to time series analysis, and it may be expected that the relative ease with which such analysis can be undertaken once an appropriate data-storage system is fully established would be a powerful stimulus.

The use of data for time series analysis may take different forms in analytical practice. In addition to applying the analysis directly to the primary data, experimentation with derived and secondary measures and with the construction of indicators and synthetic composites may, apart from serving as "barometers," assist in the building of dynamic models and in the disentangling of their components. The underlying conceptual framework may be relatively uncomplicated and predominantly descriptive of the changes in the several variables, or it may be an attempt to relate the fluctuations in these variables more explicitly. Furthermore, these attempts may be directed to one or two variables at a time or to many more; they may trace the evolution of a given CPS series, perhaps along autocorrelative lines, or they may weave the several series into a consistent and coherent tissue representative of the dynamic character of the economy as a whole. Explorations of the CPS time series along the lines of autoregression are about to become practicable without much effort on the part of the analyst who might be interested in exploring the usefulness of lag correlation and correlogram[24] for observing the behavior of the manpower, employment, and unemployment series; also, technically more ambitious and much more complex approaches can now be considered without the risk of involvement in major and very tedious computations. Thus, the availability of high-speed computers, especially when part of an integral information-storage system, brings

---

search theory, the reduction refers to the degree of dimensionality rather than to the data. Although much of the work in pattern recognition has been undertaken with a view to visual patterns — for instance, machine reading of symbols, numerals, and letters — it is in no way limited to such patterns. In the present context, reference is made to pattern recognition and search theory because of their great potential in developing methods and strategies that might make it possible to extract for intensive study the most promising conditions or alternatives from among the enormous number found in the analysis of CPS information. An example would be the lumping of conditions which are similar into large blocks of indifference classes (see, for instance, Van Court Hare, Jr., *in* David B. Hertz and Roger P. Eddison, editors, *Progress in Operations Research,* New York, Wiley, 1964, Vol. 2, p. 147). Recognition or search in the present context might refer to patterns ranging from taxonomic properties of labor force groups and subgroups to families of gross change patterns or time series of employed and unemployed by industry, sex, occupation, and so forth.

[23] See, for instance, Jon Cunnyngham, *The Spectral Analysis of Economic Time Series,* U.S. Bureau of the Census, Working Paper No. 14, 1963; from a more advanced and comprehensive point of view, see C. W. J. Granger and M. Hatanaka, *Spectral Analysis of Economic Time Series,* Princeton, Princeton University Press, 1965.

[24] By "correlogram" is meant the graph of the value of autocorrelation coefficients for different "lags" of one and the same series. (See H. Wold, *A Study in the Analysis of Stationary Time Series,* Uppsala: Almquist & Wiksells, 1938, which uses the term for the first time in statistical literature.) As to serial and autocorrelation, the terminology is not uniform; in the present context the correlation of successive terms with each other of one and the same time series will be called autocorrelation. This correlation is measured by the first, second, . . . and nth order autocorrelation coefficient according to whether the original series is correlated with the same series lagged by one, two, or *n* intervals, for instance, months.

The analyst interested in studying the formal aspects of the dynamic character of one and the same series will do well to avail himself also of the tool of spectral analysis. Where two or more series are to be correlated with each other, this is referred to as serial correlation. Two or more series can be serially correlated in terms of their "contemporary" members or after introducing various lag patterns. The latter procedure is interesting also because of its possibilities as an instrument for the study of cyclical and other time-sequenced "contagion." With the increase in length and number of series, computational work soon reaches forbidding dimensions; but use of high-speed computing equipment, where suitable, brings the problem into the realm of practicable size.

The computational as well as the theoretical aspects of auto- and serial correlations are complex, and the discussion is beyond the level of simple statistical language; to form a general impression, see, e.g., the still pertinent work by H. T. Davis, *The Analysis of Economic Time Series,* Bloomington, Indiana, Principia Press, 1941; see also, especially because of its clear exposition concerning the affinity of the computational aspects and matrix algebra for which high-speed computers are eminently suitable, J. Johnston, *Econometric Methods,* New York, McGraw-Hill, 1963, and references on p. 199.

For a compact and advanced treatment that includes ample bibliographic references, see E. J. Hannan, *Time Series Analysis,* London, Methuen, 1960.

stochastic-process analysis within the realm of the computationwise unsophisticated analyst. Beyond classical stochastic models, the analyst can now apply to the CPS data the new and powerful, although as yet little-tried-out spectral analysis. From the technical point of view, spectral analysis can be looked at as a combination of autocorrelation and so-called harmonic analysis, i.e., decomposition of a time series into component periodic terms. As the diagram depicting the behavior of the autocorrelation coefficients is referred to as "correlogram," so the graph of the intensity of frequency resulting from the harmonic analysis of time series plotted against the time period in, say, years is called a periodogram.[25]

More tempting even than the study of spectral functions of various manpower series would be the observation of their cross-spectra which are suitable for the description of lead-lag patterns between time series, in terms of their timing and interaction.[26] Expansion of the use of such analytical patterns to include "contagion" in terms of geographic categories would also appear within the scope of the enterprising analyst. Only as an illustration, it is mentioned in passing that, for the first time, it will be possible to obtain autovariance-covariance matrices[27] of reasonable size as a byproduct of the data-retrieval process itself; such matrices are of considerable importance, among others, in time series analysis of the just-mentioned kind.

More specifically, demand has been developing for data to be used in the construction of promising cyclical indicators in the form of composite measures of multiple jobholding, of part-time employment, or of hard-core unemployment. Data on multiple jobholders are now being produced within the framework of CPS in response to supplemental surveys only;[28] however, it is expected that the gathering of this information is going to be continued and data may be accruing which could be used in the construction of experimental indicators of change of a more nearly structural and longer range nature. Concerning hard-core unemployment as a criterion of change, much can be said in favor of refining conventional employment and unemployment measures to reflect varying levels of labor force attachment for selected subgroups of the population. It would seem reasonable that the less the labor force attachment the larger the influence of factors other than overall economic conditions which exert their influence on the employment or unemployment totals and subtotals. The crucial analytical problem here, as in similar situations, is the development of a standard of comparison — in this instance a standard hard-core unemployment concept or level. And it is here that the wealth of CPS information, especially when available in a format suitable for time series analysis, operates as a powerful stimulus for research into not only a level of unemployment but also into the entire continuum of varying "propensities to have a job," as intimated by the various labor force attachment figures for categories on not-too-high levels of aggregation.

Also in this class of analytical effort belongs the quantitative study of warning systems[29] to presage changes in employment levels for given subgroups of the labor force by age, sex, industry, and the like. Here again the abundance of CPS information makes it very tempting to subject the time series of components and labor force groups and subgroups, on low levels of aggregation, to the kind of analysis that other economic and business data have been exposed to for some time; experiences acquired elsewhere may be profitably carried over into the manpower-analysis field, including leading indicators and diffusion indexes.[30]

---

[25] See, e.g., Gerhard Tintner, ***Econometrics,*** New York, Wiley, 1952, and early literature there quoted. For a noted application of periodogram analysis to economic data, i.e., the 300-odd-year-long series of wheat prices, see Sir W. H. Beveridge, "Wheat Prices and Rainfall in Western Europe," ***Journal of the Royal Statistical Society,*** Vol. 85, 1922, pp. 412 ff. See also M. G. Kendall, ***The Advanced Theory of Statistics,*** London, Griffin, 1946, Vol. 2.

[26] See Cunnyngham, *op. cit.,* p. 38, and Chapter 4.

[27] A covariance matrix, sometimes called a variance-covariance matrix, is a square arrangement of rows and an equal number of columns summarizing the covariances between the several variables in a multivariate situation. Thus, if there were 10 variables involved, the first element in, say, the second row of the arrangement or tableau called a matrix would be the covariance between 1 and 2; the fourth element in the sixth row, the covariance between the fourth and the sixth variable, etc. The covariances on the diagonal, i.e., of the first element in row 1, of the second one in row 2, etc., in other words the "covariance" of a given variable "with itself," is its variance; hence, the occasionally used name, variance-covariance matrix. The covariance is a measure of joint variability, and it is defined as the sum of the products of the deviations of two variables about their respective means; the variance being an analogous measure referring to one and the same variable only, therefore, involves the products of deviations of two identical variables, i.e., the squares of the deviations about the mean of that variable. (For a nontechnical presentaion, see P. H. DuBois, ***Multivariate Correlational Analysis,*** New York, Harper, 1957.)

The covariance matrix is a useful computational device in multivariate analysis for the characterization and treatment of multivariate distributions, e.g., in more complex time series analysis, in factor and component analysis, and in input-output analysis. (See, e.g., A. S. Goldberger, ***Econometric Theory,*** New York, Wiley, 1965, for a discussion from an economist's point of view.)

[28] 1960, 1961, and 1963; see also Special Labor Force Reports Nos. 9, 18, 29, and 39.

[29] See Harold Goldstein, "BLS Occupational Trend Projections: An Appraisal," ***Monthly Labor Review,*** October 1963, pp. 1135 ff.

[30] "Diffusion index" is the name given to the ratio of the number of rising time series to all time series in a particular group of series during successive intervals. Its analytical significance is the tendency of this measure to "lead" by as much as one-half year the corresponding group of series which it describes. It thus provides "an early warning system of economic intelligence." (Julius Shiskin, ***Signals of Recession and Recovery,*** New York, National Bureau of Economic

(Continued on page 31)

The analytical topics listed above refer to dynamics in the large. Although their history within the framework of economic and social science is a long and respectable one, important problems have been appearing, some of them only recently, which raise analytical issues in what might be called dynamics in the small, or micro dynamics.[31] Since quantitative information of the kind required for the analysis of micro dynamics of the labor force and of manpower in general has been difficult to come by, efforts along this line have been much less frequent than in the economic dynamics on high levels of aggregation. Yet some of the most significant and challenging questions can hardly be answered without recourse to the dynamics on low levels of aggregation.

Among the many analytical challenges in this category is the entire family of mobility problems. Foremost among CPS sources for mobility studies — mobility in the broad sense of the word — are the gross change or gross flows tabulations. As earlier mentioned, gross change statistics are beset with difficulties and complexities in addition to the usual ones. This is one of the reasons why their publication was stopped in 1953. However, actual tabulation has been continued on a monthly basis so that by now a respectable body of data is available covering a period of many years, going back to 1948. Also as mentioned earlier, gross changes refer to changes from one month to the next in employment status of individuals or, for that matter, in other characteristics pertaining to individuals. To the extent that the data reflect changes in the status of individuals rather than of entire aggregates, one is here dealing with micro information. Other sources of similar statistical information, although referring to different characteristics, are the several supplemental surveys, for instance, the ones on job mobility,[32] on work history,[33] and, to some extent, the supplemental surveys on job tenure and on work experience.[34] The basic CPS schedule contains, of course, a host of micro data and could "in principle" be linked from month to month for identical individuals. Although such information might be obtainable if data on available microtapes were to be retabulated, the only actual tabulations are those included in the gross change tables.

Among the mobility problems for analytical exploration alluded to most often explicitly or by implication are those referring to inter-labor-force mobility; to intra-labor-force mobility primarily as between occupations, industries, and jobs; and to geographic mobility. Of particular interest seem to be questions as to what socioeconomic characteristics are associated with mobility; as to the degree of mobility; and where mobility causes a worker to move through several "states," "industries," and the like, as to the "original point of departure" of those who finally became unemployed. Frequently the underlying interest is more general, resulting from questions such as how people adjust to changes affecting their labor force status or job, and what the characteristic gross flows patterns are for specific subgroups of the labor force or for those outside the labor force because apparently not much further insight can be gained from the study of only the largest aggregates. Similarly, the analysis of structural unemployment often requires micro information on mobility, especially where "last" employment of a now unemployed person may obscure or vitiate the actual pattern unless consideration of mobility is explicitly introduced. Also problems of mobility, although complementary ones, are those pertaining to relative stability of the labor force and its major components.

The study of inter-labor-force mobility, that is, of the movers into and out from the labor force, was briefly discussed at an earlier point. Here again analytical conclusions will often be based on comparisons with a norm; hence, descriptive measures will be helpful of average, typical, most frequent, or otherwise characteristic behavior. Once the data are accessible, it would be easy enough, though tedious, to establish such behavior patterns for specific groupings as to age, labor force status, industry, and color.

As to intra-labor-force mobility, a move or shift of major importance is the passage of an individual from employment into unemployment status or vice versa. There is ample information available currently, but only for comparing employment status in two neighboring months at a time. Although this is no serious limitation where one intends to reconstruct time series describing the changing mobility picture, month by month, the restriction to pairs of neighboring months does prevent longitudinal studies of a cohort type where individuals would have to be traced through a number of successive months. In other words, the available information makes it possible to utilize unbiased estimates though at some loss of efficiency, for the reconstruction of mobility patterns over time in terms of a chain of sliding successive bimonthly links, but it is now not feasible with the help of gross change information to ascertain, say, the industry of a person's employment beyond that of the first month in each of the two-month links. Even if the length of these links were to be expanded by collating

---

Research, 1961; see also Geoffrey H. Moore, editor, *Business Cycle Indicators,* Vol. I, *Contributions to the Analysis of Current Business Conditions,* a study by the National Bureau of Economic Research, Princeton, Princeton University Press, 1961.)

[31] See, for instance, William S. Vickrey, *Microstatics;* and *Metastatics and Macroeconomics,* New York: Harcourt, Brace and World, 1964 (both works).

[32] 1955 and 1961; see also *Current Population Reports,* Series P-50, No. 70, and Special Labor Force Report No. 35.

[33] See Special Labor Force Report No. 37.

[34] See *Current Population Reports,* Series P-50, No. 36, 1951; Special Labor Force Report No. 36, 1963; and Special Labor Force Reports, *Work Experience of the Population,* Nos. 11, 1960; 19, 1961; 25, 1962; and 38, 1963.

the micro information for a sample of identical individuals, the rotation feature of CPS would limit the maximum length of such links to 15 months, with an 8-month blank in the middle. However, such an extension may, in periods of rapid job shifts between original employment and final unemployment, improve the chances considerably for identifying the proper industry or occupation of "origin." To undertake a farther researching analysis would require supplemental questions aimed at reconstructing work histories or possibly the matching of CPS with Social Security Administration records where this is feasible. However, even when limited to bimonthly flow patterns, which for many purposes can be pieced together without any serious loss of validity, one can trace a great many changes. There is thus room for a great deal of descriptive and analytical effort directed to the study of mobility patterns of rather small aggregates.

For cross section type of analysis, a unique source of information on job mobility is the February 1962 Supplemental Survey which is summarized in Special Labor Force Report No. 35; this report contains several detailed tables which describe, for 1961, job-shift patterns by age, sex, occupation, industry group, and marital status; it also contains a rudimentary job-shift matrix by "occupations left" and "new occupation"; in other words, the tables extend the two-month link of the gross flows tables to an entire year, but for one calendar year only, namely, 1961. Again, there were considerably more tabulations made than were published in the just-mentioned Special Labor Force Report.[35]

Also pertaining to the mobility problem are some of the data produced by the supplemental survey on "length of time on present job." Only two such surveys have been undertaken (in January 1951 and 1963); they are summarized in *Current Population Report,* P-50, No. 36, and in Special Labor Force Report No. 36. The tabulations can be used to reconstruct job-life for jobs now held by age of worker, by detailed industry group, by occupation, for women by marital status and age, etc. Subjecting, therefore, the various tables to appropriate analysis might yield insight into the differential patterns of job-life distribution which otherwise are not easy to come by. To the extent that sufficient information would be available in the various table cells, relative stability and instability in job tenure could be explored beyond the now available studies by industry, occupation, sex, broad age group, etc., in order to let the empirical data suggest what factors are likely to account for length of job-life, which in turn is inversely related to job mobility.

Among the several mobilities, one which has commanded attention for some time is geographic mobility. The relation between internal migration and labor force characteristics has been of obvious analytical interest; more recently particular attention has been given to questions such as the suspected inverse correlation between mobility and length of unemployment, and the related one of a positive association of mobility and income; thus geographic mobility has come to be considered one of the means of adjustment of the labor force to a changing labor market, and "propensity to move" one of the important characteristics of the various components and subcategories of the labor force.

CPS regularly includes geographic mobility information but not as a part of the basic monthly CPS schedule; questions on mobility have been added to the basic schedule once a year,[36] and they aim at ascertaining whether a particular person had moved to the present address within the last year, and if so whether he had moved within the same county, within the same state, from another state, or from a foreign country.[37] Therefore, tabulations of the information can be made according to whether the move, if any, during the period covering the 12 months from March to March took place between different addresses of the same county (movers) or between different counties (migrants); the latter group is subdivided in the tabulations into "within a state," "between different but contiguous states," and "between noncontiguous states"; schedules referring to persons who during the 12-month period under consideration had moved to their present address from abroad are also separated in the tabulation.

This part of CPS was not included in the transfer to BLS under the 1959 agreement with the Census Bureau; and although microtape data are "in principle" still available for the past few years, the primary source data going back in time are the tables published in the CPS Series P-20 and their predecessors, containing mobility information as far back as 1940.[38]

Although these published tabulations are quite extensive, the actually processed materials have, of course, exceeded the content of the published tables.[39]

The analyst will often be interested in the kind of information that can be reconstructed over the past.

---

[35] The information is unique in several respects, in spite of some serious limitations, in that it contains, for instance, tabulations by earnings and by rather detailed occupational and industrial groupings, and by reason for leaving jobs; less than half of the tabulations were found suitable for publication.

[36] The so-called mobility supplement to the Current Population Survey, usually for March.

[37] Questions 34 through 36 on recent schedules.

[38] See Labor Force Memorandum No. 1; CPS Series P-5, Reports Nos. 5, 6, 8, 11, 14, 15, and 24; and CPS Series P-20, Reports Nos. 4, 14, 22, 28, 36, 39, 47, 49, 57, 61, 67, 73, 82, 85, 104, 106, 111, 113, 118, and 127.

[39] For instance, "mobility status" and "type of mobility" tabulations were made of white-collar workers and manual workers in greater detail than shown in the CPS tables; information was also tabulated for service workers, separately by private household and other service workers; for selected "work experience" groups; and for six income intervals.

Mobility data from many of the earlier CPS reports include information on labor force status and on employment status, covering the entire period from 1940 to date.[40] For other than labor force and employment status items, the information by mobility characteristics is spotty, especially for the period before 1959. However, the cross section type of comparison over time might well be attempted, in spite of the absence of true continuity of the data; for instance, mobility could be studied by broad occupational groups.[41]

Recapitulating, it was found that although the gross change data are the only ones regularly tabulated, a great deal of supporting data have been produced by intermittent or one-time supplemental surveys.

In weaving the various materials skillfully and patiently together, considerable additional knowledge might be acquired through appropriate analyses of such aspects of the manpower problem as the relation between mobility and unemployment, between length of holding a particular job and industry or occupation and age; about occupational and industrial mobility in general and its changes over time; about the pattern over time of entries into and exits from the labor force by a variety of characteristics of the movers; and about the trends of the inter-labor-force mobility of the unemployed.

Although the mobility data suffer from a variety of shortcomings, they do yield an enormous mass of raw material which might usefully be applied to pilot mobility studies in order to develop the now missing analytical framework for such research. Once one realizes the staggering number of possible patterns existing in a monthly series, even if it covers but a few years, he will be impelled to explore stochastic-process models and similar multivariate devices with a view to their descriptive suitability before even attempting analysis proper.

Global Data Schemes and Models

Finally, CPS data use will be explored for the analysis of a few problems which, although substantive in nature, are on the borderline of methodology. The kinds of problems here involved fall into the area of model building and of comparable but descriptive or idiographic schemes and systems.

The concept of large-scale systems for application in the presentation and the analysis of data has been attracting and fascinating social scientists for a long time. But only recently can there be discovered a similar development in the field of labor economics in general and in manpower analysis in particular.

Global systems of a comprehensive nature describing the national economy or some of its broad sectors represent the ultimate of stress on their empirical data foundations. No single data source, even a detailed census, is usually capable of supporting an overall and all-embracing model or scheme without the help from other data sources and without auxiliary constructs. It would be unrealistic, therefore, to expect a single data-producing system, be it as comprehensive as CPS, to provide the necessary information in the field of manpower. On the other hand, CPS is the next best source to a census, and it has the great advantage of being a monthly survey and a great deal more flexible, considering the availability of supplemental surveys. Also, it has been a vehicle for some more specialized or delicate inquiries than can be or have been undertaken through censuses.

Among global problems, the relation between labor force and economic structure has been of obvious interest to economists and others for some time. Most recently this concern has been restated by the President's Committee To Appraise Employment and Unemployment Statistics.[42]

The subject can be broached in its simplest form using such devices as the analysis of concomitances of gross national product and its components with labor force totals and their components. Here CPS data are of some interest inasmuch as they can be applied to census and similar benchmark data in order to carry them forward and keep them current or to interpolate between benchmark dates.

A technically not dissimilar undertaking would be the attempt to trace, for the economy at large, the changing sociodemographic characteristics of various labor-force-status groups. From here it is only a short step to the vast area of occupational structure studies, which in one form or another have become of growing significance for the solution of manpower utilization problems, be they on national, local, or on individual company levels. Since one of the objectives of occupational structure analysis is to gain insight into skill mixes and requirements of particular jobs and occupations, hence industries and economic sectors, some of the more recent advances in micro analysis in the so-called management science and operations research may be of interest to the user of CPS. True, data of the kind collected through CPS do not compare in detail with the data accumulating usually in, say, personnel departments. On the other hand, their scope and coverage, and the length of time over which comparable data are now available, open avenues not accessible to those who would restrict their analysis to materials in personnel files — even of the largest companies or the federal government.

---

[40] Prior to 1963, information on full- and part-time employment is available for the period 1958-1961 only, and on work experience for 1959-1961; "class of worker" statistics were published for 1941-1945, 1953-1955, and 1958-1962; industry group data, for 1940-1945, 1950-1951, and 1953-1954. For occupational information, the data cover the years 1941-1951, 1955-1956, 1958-1959, and 1961-1962. For four past surveys, the number of weeks worked was published; for two, the number of weeks unemployed; and for five, income, with no cross-tabulation by labor force characteristics. Annual data are generally available as listed above for the years after 1947-1948 only.

[41] On qualification of data, see in particular the recent publication by Henry S. Shryock, *Population Mobility Within the United States*, Chicago, University of Chicago, 1964.

[42] *Measuring Employment and Unemployment*, 1962, p. 72.

As pointed out at the outset of this section, the kind of problem here discussed falls into the border area between substantive and methodological analysis. Much of the work to be undertaken with CPS data within this context would, therefore, have to be of an exploratory nature.

For instance, to discover suggestive concomitances, CPS-based time series for a great number of subgroups, subtotals, and estimates referring to tabulation "cells" on various levels of aggregation could be projected against or superimposed over economic time series representing cyclical indicators or other symptomatic quantities.[43] Such superimposition may convey a fairly good picture of the time flow of manpower characteristics and variables relative to other aspects of the economy; it would also help in comprehending interrelations between various series that can be grasped directly or by the use of straightforward descriptive measures, for example, measures involving lead and lag patterns, or "conformity" and nonconformity; furthermore, the data can be used as empirical evidence on the presence or absence of invariances over time and, therefore, increase or decrease the analyst's confidence in the stability of observed patterns where forecasts or prognoses are involved. In addition to this kind of analysis of CPS data, the dynamics of more complex interrelationships can be made the objective of the inquiry. An example is the more nearly classical type of market analysis which only recently has been capturing the full attention of labor and manpower economists. Here interest in CPS data seems to be developing along cross section as well as longitudinal lines of exploration, the emphasis being on the dynamic aspect of the labor market structure. The resulting combination of information requirements and the need for data on lowest level of aggregation in order to study the interrelation of household and individual characteristics with the supply of and the demand for labor, as well as the emphasis on the dynamic and process aspect of these interrelationships, make CPS a particularly noteworthy source of information. And to the extent that mass data with nationwide coverage are desired as a foundation for the analysis, CPS would appear to be the only promising source available at this time.

Put differently, although the approaches, models, and analytical techniques span the wide area from cross section to time series analysis, or from autocorrelation of discrete and distinct series to serial correlation or more intricate interaction patterns, the information requirements are similar: empirical data are needed on lowest levels of aggregation and available for frequent, successive points in time. CPS data, therefore, would be in great demand for such analysis, especially if available on micro levels. This interest in micro levels is further heightened where CPS is being used as a sampling frame for, say, intensive interviews of a subsample of CPS households so as to implement information obtained through regular CPS.

Methodologically related are the analytically oriented data uses for the study of the occupational structure. Although census data provide here more complete and detailed information, growing interest in CPS can be traced to two circumstances: the increasing attention to problems of a changing occupational and skill mix; and, second, the realization that some highly relevant characteristics of the micro elements to be studied are themselves dynamic in nature and therefore require for their identification longitudinal-type information. Examples are the need for recognizing individuals with different propensities to move, and their particular geographic, job, industry, and other mobility patterns; those with various labor force attachment and employment-unemployment sequence patterns, and so on. Finally, there arises in the distance, although in dim and vague shape, the problem of a comprehensive manpower resources accounting system, incorporating broad skill levels and occupational affinities as well as the traditional socioeconomic and demographic properties.

On micro levels the problem is not new, and various approaches have been contemplated and devised. Within the context of governmental efforts — especially defense-oriented efforts — the problem has been considered in its more generalized aspects, and it occupied a prominent place as early as Project SCOOP[44] under the label of Personnel Assignment Problem. Model building, optimization, maximization or minimization, and the finding of other solutions often proved, after an initial period of coping with the problem, to be less of an obstacle than the seemingly simple question of how and where to obtain the empirical data to be fed into the model. Related approaches to what, ultimately, is an occupational structure problem can be found within the framework of interindustry analysis. Although sometimes forgotten, it is a historical fact that among the main objectives of early thoughts along the lines of interindustry analysis there were the problems of labor force and manpower utilization.[45] Since these early days, continuing attempts have appeared in the literature to accommodate manpower and employment problems directly or indirectly with the help of the input-output scaffolding; and a good deal of thought has been given to the resulting technical aspects, so much so, that two of the specialized input-output bibliographies, one published by the Operations Research Office of the Johns Hopkins University and the other by the United Nations for the period 1955 through 1960, devote entire sections to the special problem of manpower.[46]

---

[43] A conceptual example for a synthetic quantity of this kind is the "reference cycle," treated by Arthur F. Burns and Wesley C. Mitchell, ***Measuring Business Cycles,*** Studies in Business Cycles, No. 2, New York, National Bureau of Economic Research, 1946.

[44] Scientific Computation of Optimum Programs; see especially D. F. Votaw, and A. Orden, ***Symposium on Linear Inequalities and Programming,*** Washington, Headquarters, U.S. Air Force, 1952, pp. 155 ff.

[45] See, for instance, the diagram summarizing the Basic Analytic Framework, p. 166, and the brief section on interindustry relations, p. 187, *in* Jerome Cornfield, W. Duane Evans, and Marvin Hoffenberg, "Full Employment Patterns, 1950," ***Monthly Labor Review,*** February and March 1947.

There thus appears to be no shortage of ideas and analytical interests, but there is a bottleneck on the data-input side. Therefore, CPS data should most certainly be explored, especially where information is needed reflecting changes over time, and on relatively low levels of aggregation. As earlier mentioned, the demand on empirical information of such comprehensive and ambitious undertakings puts a great stress on any empirical foundation, and no potentially useful source of information should therefore be left unexplored. The development of frames and scaffoldings such as the ones just touched upon, be they in the form of a personnel classification tableau arranging persons alongside jobs in terms of appropriate skills, occupational affinities, and expected productivity,[47] or in the form of an interindustry matrix kind, is therefore of considerable analytical possibility and promise, even without any commitment to some of the more ambitious uses for which the input-output framework may sometimes have been intended. An analytically meaningful and further manipulable multivariate frame or matrix of observational data is itself a first, important step forward; it is also important as a powerful and systematic data-specification and programming device. The descriptive function alone of such an idiographic scheme, modest as it may seem, is of considerable practical significance for real understanding of some of the complexities of the manpower problem because it enables the analyst to see behavior patterns on lower levels of aggregation than heretofore possible.

Such occupational idiograms can therefore help the analyst in perceiving the various component sectors of the structure and their changes in terms of, say, age, sex, and educational achievement mix; what the age, sex, educational achievement, etc., requirements are that would be generated by exogenous changes in the occupational structure; and, on the assumption that occupational distribution is closely related to distribution of output among the various industries and sectors of the economy, what skill and manpower requirements would be associated with anticipated changes in the output structure.

The strength of the approach, therefore, need not be in its general theoretical power but could result from its taxonomic accomplishment and the ensuing suggestiveness of observable patterns, relationships, and even of simple distributions of the desired characteristics over relevant subgroups and subcategories of the labor force — and all that on low and flexible levels of aggregation. For instance, an appropriate taxonomy would help in locating manpower sectors of actual or threatening oversupply, in identifying segments where shortages exist or are likely to arise, and in developing manpower allocation schemes where desirable; it could also be instrumental in establishing, within the information structure, areas in which serious gaps or other deficiencies appear in the available data supply. It would enable the analyst to observe changes in the relationship between levels of unemployment, in unemployment mix,[48] in interoccupational differentials of socioeconomic and demographic characteristics and, to the extent that the data permit, in the concentration of unemployment by industry, type of worker, and other relevant properties.

The emphasis thus is on the supply of empirical data which assumes the role of an input into a descriptive manpower account with considerably more analytical potential than has as yet been fully grasped. In this connection, CPS information potential deserves further exploitation, especially where change over time is of primary importance. An example is the model recently developed by R. J. Hollister in connection with manpower forecasting.[49] In the description of this model, Hollister points to the great need for analyzing data on change rather than on levels of occupational distribution in order to form an opinion about the character of the all-important skill coefficients.[50] The dynamics of these skill coefficients, caused by changes in relative supply and prices of skills and by changes in technology, assume a central place in the model; they also are the determinants of the success or failure of manpower forecasts under all but the most trivial assumptions. To the extent that the occupational distribution may be assumed to reflect, even though with some distortion, the distribution of skills, judicious experimentation with CPS data might reveal insights into the dynamics of the skill composition of the population. By extending the occupation-by-industry tableau to include other relevant dimensions describable with CPS data, the broad outlines of the changing multivariate manpower profile might become visible.

---

[46] Vera Riley and Robert Loring Allen, *Inter-Industry Economic Studies,* Baltimore, 1955, with separate sections on manpower analysis and projections, manpower requirements, labor supply models and labor mobility, occupational structure, and area labor requirements; and Charlotte E. Taskier, *Input-Output Bibliography, 1955-1960,* New York, 1961, with a special section on manpower studies.

[47] See, for instance, Votaw and Orden, *op. cit.,* p. 186.

[48] John P. Dunlop's terminology, "class versus mass unemployment."

[49] "The Economics of Manpower Forecasting," *International Labour Review,* April 1964, p. 371.

[50] In analogy to input-output models, the skill coefficients a_{ij} are defined as the number of units of skill i per output of industry j in $L_{ij} = \Sigma a_{ij} a_j$, the total input of skill i.

Suitability of the Data for Methodological Analysis

The so-far-discussed utilization of CPS data refers to direct analysis of manpower and labor force characteristics and behavior. In addition to such substantive analysis, there is also methodology-oriented analysis; hence, there is the question of analytical exploitation of CPS data also along methodological lines. Use of data for the purpose of improving methodology is, of course, an important undertaking in its own right. Only in passing should it be pointed out that, if successful, much of the methodological effort may in turn result in extension of the information basis or its equivalent — hence, in an even higher analytical potential of CPS data.

Among the attractions of CPS from the methodological point of view are its excellent technical level, its impressive time coverage, the frequency of the survey, the Census Bureau's great and persistent interest in the survey's quality control, and the amount of systematically accumulated information on survey and sampling experience over a long period. Most of the potential uses of CPS data with a view to methodology will, therefore, not be novel to those intimately connected with the operation of the survey program. In spite of this, it is felt that because of limitations of staff, time, and resources, this potential has not been fully exploited and therefore does represent real analytical potential in the pragmatic sense of the term. The condensed listing of illustrative problems in this category is divided, for purposes of presentation, into four sections, obviously without claim to completeness:

1. Survey-technical aspects;
2. Taxonomy, concepts, and definitions;
3. Linking, matching, and collating the data;
4. Decomposition of time series and related techniques.

Survey-Technical Problems

A survey operation of the caliber and difficulty of CPS generates a host of methodological problems, frequently as byproducts of the operations themselves. It is not intended here to give even a summary account of such problems but rather to highlight some methodological problems which suggest the use of CPS data for their solution. In other words, the emphasis is on CPS data, and on their methodology-oriented exploitation. Many of the most interesting and significant problems must, therefore, be omitted because their solution requires efforts other than the analytical use of available CPS statistics.

With this limitation in mind, examples of data uses for survey-technical problems can be found in the researches on hand of CPS micro data, in noninterview imputation and substitution effects (for instance, the influence of substituting variously defined sample segments), or in the analysis of the effect of response errors on multivariate cross-tabulations and measures of association.

All these problems raise the question of nonsampling error and how to cope with it. The interest of the manpower and labor force analysts in these questions, however, need not be purely methodological because the answers to these questions may have considerable influence on their interpretation of the substantive information. Because of this close connection between method and substance, there is not necessarily a single and simple answer to the question of the importance of a nonsampling error and how to correct for it. Ultimately the analyst must be the judge of the meaning and applicability of the data.

Taxonomic, Conceptual, and Definitional Questions

Concerning taxonomy, concepts, and definitions, various ways of forming labor-force-status categories have been suggested at one time or another;[51] for categories and definitions based on class intervals, the problem of forming taxonomically optimal groups is nearly always present. Examples are tabulations for continuous variables such as hours of work, weeks of looking for a job, number of children, age in years, and the like. Variations in definitions, for instance, for youth, partially employed, and long-term unemployed may have quite unexpected statistical consequences; and appropriate exploratory manipulation of the data would seem highly desirable. In the absence of a theoretical framework within which to optimize taxonomic classifications, approaches might be considered ranging from simple trial and error to the much more complex and powerful attempts of pattern search (see footnote number 22, page 28) and to statistical experimentation along the lines of variance and covariance analysis, thus developing *ad hoc* principles of optimum classification as the analysis proceeds.

A related and also essential taxonomic question is the one as to which variables among a larger number are the relevant ones or the productive ones in terms of the envisaged analysis. The problem is particularly to the point because CPS is a sample survey, and greatest economy with available observations or degrees of freedom is always a good rule of thumb. An example is the question of income as a classifier compared with employment status. Inquiries may here proceed along the line of classical component analysis (see footnote number 19, page 26), of discriminate analysis (see footnote number 18, page 24), and finally of simulation[52] to observe which, if any, differences result from taxonomic variations.

---

[51] See, for instance, Gertrude Bancroft, "Current Unemployment Statistics of the Census Bureau and Some Alternatives," *in The Measurement and Behavior of Unemployment,* a Conference of the Universities-National Bureau of Economic Research Committee, Princeton, Princeton University Press, 1957, pp. 63-119.

Linking, Matching, and Collating the Data

Of particular interest to the substantively oriented analyst are the methodological problems inherent in linking, matching, and collating data — the very hub of longitudinal studies.

Problems in this group include those related to gross change or gross flows measures already mentioned earlier — methodological aspects of the matching micro records from several statistical sources whether aiming at CPS records only (for instance, supplemental surveys to be matched with basic CPS-FOSDIC schedules and in turn with basic CPS control cards) or at CPS records in conjunction with other sources (for instance, Internal Revenue Service statistics or Social Security Administration records). Finally, the various statistical problems inherent in the panel type of operations extending over longer periods than two consecutive months also belong to this group of problems. Analytical efforts would here have to explore avenues in the direction of a study of characteristics of movers, the susceptibility to conditioning of the various survey items, their possible time pattern, and the like.

Decomposition of Time Series and Related Techniques

In the fourth group of problems, there are the technical aspects of time series analysis which are here of signal analytical import in view of the very nature of statistical data-storage and -retrieval systems in terms of their expected capabilities in reconstructing time-sequenced chains of data on lowest levels of aggregation.

An example of analytical patterns which offer themselves at the outset are the various techniques of eliminating seasonal variations. Recent progress notwithstanding, there remain a great many difficult unresolved problems which point toward the possibility of experimenting with the data. For instance, some of the observed seasonal patterns, especially those of the unemployment series, seem to display multiplicative as well as additive effects,[53] whereas multiplicative adjustments alone would seem to be satisfactory in most other situations. Also, this combined multiplicative and additive effect seems to be more pronounced during some months than during others. Here various possibilities offer themselves: experimentation with both adjustments simultaneously, or sequentially — in both directions, i.e., additive before multiplicative and conversely.

It would therefore seem in line with the recommendation of the President's Committee To Appraise Employment and Unemployment Statistics[54] if additional inquiries were to be made regarding the construction of explicit models for the decomposition of time series, applicable to the great masses of manpower data now being produced and stored.

Still within the general conceptual framework which conceives of the particular series as self-contained evolutionary sequences of data would be the investigation of the series using the much more flexible tool of spectral analysis.[55] Here the development, comparison, and further analytical condensation or reduction of the large number of spectral profiles might shed light on the applicability of new approaches toward the discovery of time patterns; in addition, such study of individual spectra simultaneously may perhaps disclose some of the aspects of what might be construed to be essentially a process of contagion or propagation throughout the economic structure generally, and the labor market in particular.[56]

Finally, there is the already touched upon possibility of examining the applicability of stochastic-process models — visualizing the CPS series as interdependent components of an entire system moving along a common time axis. Gross flows type of data should be contrasted from a methods point of view with more truly longitudinal data[57] in order to determine their respective advantages and disadvantages, and analytical weakness and strength when used for process analysis.

Briefly, then, it may be concluded that there is as yet promising unused potential contained in the labor force and manpower component of the CPS. From the analyst's point of view, the usefulness of the data is directly proportional to the ease with which it can be retrieved, the flexibility of the form in which it reaches the analyst's shop, and, last but not least, the extent of and access to a data inventory suitable for use.

---

[52] The term "simulation" is here used to describe the use of electronic data-processing equipment in the modelization of at least partially observable processes and structures and in the study of their behavior and reaction under conditions of change. In general, it would, of course, be quite possible to create purely synthetic models without any relation to things observable. From the point of view of electronic data processing, simulation could perhaps be described as the use of high-speed computers for the purpose of experimentation rather than of computing. The method of simulation is rapidly expanding, and references to it can be found in nearly any recent treatment of operations research. Among earlier sources, see D. H. Orcutt, M. Greenberg, J. Korbel, and Alice M. Rivlin, *Micro Analysis of Socio-Economic Systems,* New York, Harper, 1961, which contains stimulating suggestions; and *Report of the Second System Simulation Symposium, February 1959,* American Institute of Industrial Engineers, 1960, which describes a variety of not-too-technical aspects.

[53] Put differently, the multiplicative model assumes that the seasonal effect is proportional to the level of the series, the additive model that this effect is independent of the level.

[54] *Measuring Employment and Unemployment,* 1962, especially Chapter VI.

[55] See, for instance, Jon Cunnyngham, *The Spectral Analysis of Economic Time Series,* U.S. Bureau of the Census, Working Paper No. 14, 1963, which contains a good bibliography on pp. 87 ff.

[56] See, however, also the cautionary conclusions by Harry M. Rosenblatt *in Spectral Analysis and Parametric Methods for Seasonal Adjustment of Economic Time Series,* American Statistical Association, 1963 Proceedings of the Business and Economic Statistics Section Annual Meeting.

[57] For the use of the Social Security Administration's continuing work-history sample materials in stochastic-process analysis, see, for instance, Isadore Blumen, Marvin Kogan, and Philip J. McCarthy, *The Industrial Mobility of Labor as a Probability Process,* Ithaca, Cornell University Press, 1955.

Given reasonably favorable conditions and terms for its use, CPS information is of considerable interest to the analyst, especially where he can take advantage of the ease with which the data can be subject to aggregation on different levels, including the very lowest ones and also of the continuity of the data for time series and longitudinal analysis. Especially for micro data ordered in time, which are the raw material for longitudinal studies in the narrower sense of the word, CPS is — in spite of limitations — a unique source of information within the system of United States government statistics.

And it is along these two lines, disaggregation and assembly of micro data over time, that the analyst's desire and need for more ample information are most pronounced. The quest for disaggregation will always meet, sooner or later, with the obstacle of insufficient and unreliable information. Ultimately, therefore, expansion of sample, reduction of nonsampling error, and further improvement, if possible, in estimation procedures might help to reduce this obstacle but would never be able to remove it completely. On the other hand, it is felt there is now available "in principle" a good deal more information of this order of refinement than the prospective analyst has access to or, for that matter, knowledge of.

Concerning micro information, the analytical serviceability of the data could be greatly increased if gross flows type of statistics were made available and further expanded and if the longitudinal coverage could be stretched to cover a period of more than two adjacent months. It is here in particular that further research of a methodological nature, including the critical problem of the effect on response of panel operations, would appear most promising; however, the possibility of actual expansion of gross flows type of information ultimately over the entire life of the rotation group should not be excluded, nor should the experimentation with different rotation patterns extending over longer than the now established periods.

Finally, there would appear the possibility of providing additional analytical raw material without actually expanding survey operations or questionnaires by collating information already being collected. Examples within the system of CPS in the broader sense would be the matching of supplemental with subsequent basic CPS schedules, and of control cards with FOSDIC schedule. Furthermore, it is only an expansion of this idea to conceive of all manpower statistics as ultimately forming one comprehensive data system. As a step in that direction, consideration should be given to matching CPS with information resulting from entirely different data sources within the government (e.g., Internal Revenue Service, or Social Security Administration) or outside the government where appropriate.

VI. Summary and Outlook

The will and the ability to exploit the analytical potential of a given body of information are predicated on the presence and interaction of many factors; among them intellectual curiosity, creative imagination, and technical skill furnish the necessary although not always sufficient ingredients. Even more complex are the circumstances which determine the route along which curiosity and ability drift and the particular direction into which they turn. Utilization of CPS manpower information is no exception.

Pressure of acute and immediate socioeconomic problems has provided much of the general impetus and motive for the study of employment and unemployment. At the same time, the close connection between the desperate urgency of what appeared to be an issue of major social, economic, and political momentum and the visionary hope for remedial action not infrequently tended to obscure the soberly analytical aspects of what had become an issue charged with intense emotions and strong impulses. It is therefore not surprising that the practical exigencies of suddenly emerging calamities have not provided an environment conducive to taking the longer view as, for instance, in the not unrelated field of demography. As in other fields, however, where attempts to cope with pathological situations sooner or later have led to the recognition of the importance of, and a growing interest in, the broader fundaments of a physiological science from which to investigate the conditions of disease; so the struggle with the catastrophic outgrowth of the major economic crises has gravitated to the gradual development of business cycle analysis and economic dynamics, and the endeavor to mitigate the unemployment problem has veered toward the much wider area of manpower analysis.

But the problem of the exploitation of CPS data is an important and fascinating one beyond the specific questions of manpower research. Analysis-oriented demand for data is not static. As the socioeconomic scene moves along, new problems enter the horizon of the researcher. Just as fashions change, so do styles of human thought. And parallel with such changes, there has come the evolution of analytical patterns of thought and inquiry, and of research techniques made possible by improvements and refinements in instrumentation, e.g., in computer technology. Evolution in the area of human thought and of analytical techniques usually brings with it, sooner or later, a corresponding expansion of the information base, especially if the trend is toward empirical research. And the new information, in turn, generates a powerful feedback which acts as a catalyst on the creation of new patterns of thought, of fresh analytical ideas, and of novel and exploratory hypotheses; if successful, these developments soon produce additional demand for data, thus continuing the endless process of alternating expansions in demand for, and in supply of, observational data.

In a highly dynamic situation such as that characteristic of the behavioral sciences during this generation, a perfect and residue-free dovetailing of data supply and analytically oriented demand for data is not expected. A surprisingly large amount of CPS information has lately been finding its way into analytical channels, but there remain major areas of unsatisfied demand, illustrating the continuing discrepancy between supply of CPS data and analytically oriented effective demand for the information. For example, it was shown earlier that among the emerging topics of major analytical interest, there is an increasing emphasis on the dynamics of manpower in one form or another, and on the path followed by individuals of specific labor force groups through the maze of taxonomic stages, as illustrated by successive labor-force-status phases. The analyst's attention has been attracted not only by the direction of the paths but also by their length, that is, the time it takes individuals to travel across the several nodes of the network or points of the road map descriptive of the manpower situation. Of particular interest are differences in patterns relating to specific manpower categories and groups, for instance, women, ethnic minorities, specific age groups, and the like. Considerable concern has also been expressed about occupational and industrial distributions, especially with a view to evaluating structural unemployment and related problems.

To come to grips with some of the more intricate economic aspects of manpower utilization, a general quest has been noticeable for additional data, for instance: data on labor force aspects of high school dropouts, the disabled, and the multiple-jobholders; data on wage rates in connection with CPS-type information; and general data of a kind that would permit analysis within the reference frame of the family and the household and its members rather than of individuals.

Although only general in nature and not exhaustive, the above examples, together with the specific cases alluded to in the preceding chapters, help to illustrate the character of the information problem; and they provide a background against which to project this question: Is the collection of additional data the answer to the unsatisfied demand for data?

On the preceding pages the point was being made that, insofar as CPS-type information is concerned, many other avenues might profitably be explored before embarking on an expansion of questionnaire and survey. Foremost among the possibilities of squeezing out more from already collected data prior to enlarging the collection of data are:

1. *Collating* information from the several surveys and schedules utilized in interviews of households in the CPS sample or its subsamples, as for some of the supplemental labor force surveys — also matching the FOSDIC schedule with control card information[1]— and collating with the basic CPS schedule the data already obtained from *ad hoc* and intermittent surveys conducted at the request of government agencies outside the framework of the labor force survey program proper, e.g., the Department of Health, Education, and Welfare and the Federal Reserve Board.
2. *Matching* the CPS schedules with information obtained from basically different data-collection sources, such as the Internal Revenue Service's tax returns or the Social Security Administration's operating statistics derived from source documents emanating from the administration of the social security system.[2]
3. *Linking,* over time, CPS information pertaining to identical members of given rotation groups in the CPS sample throughout the survey-life of the group; in other words, expanding the present gross flows data concept to include more than a pair of neighboring months.

Most of such attempts to expand the effective supply of data would require additional research and methodologically oriented explorations before usable material could be produced. Notwithstanding the possible limitations inherent in the data, the popular alternative is even less attractive: the collection of data for what at best could be qualified as nongeneralizable case studies. The solution of the problem, therefore, seems to point toward the pooling of available statistical raw materials with a view to optimizing their overall analytical usefulness in spite of the resulting need for additional manipulation and processing.

Even so, there will remain major gaps unless the CPS operation is actually enlarged and unless the number of questions on the schedule is increased. An example is the heavy demand for data disaggregated along geographical lines. Here only a substantial increase in sample size or the inclusion of particular self-representing primary sampling units would bring a noticeable degree of relief.

In summary, it would appear that — short of attempts to collect new and more data — the decision on how to deal with the unsatisfied demand for data should be made in terms of the following possibilities aimed at freeing latent analytical potential in available statistical raw material: improvement of quality through the continuing reduction of nonsampling error; some further improvement of estimation procedure if this is possible; longitudinal linking of data over several surveys along the line of pseudo-panel operations; and tying CPS data in with other major information systems, using CPS as the basis for the selection of sampling units for which the matching is to be attempted.

Whereas it is, in general, possible to meet effective demand for statistical information, provided it promises significant results and is costwise and otherwise justifiable, the reverse is not necessarily true. If available information potential is unused, it is a much more complex and delicate problem to push analytical demand up to the level of available information supply.

In essence, such stimulation of intellectual curiosity — channeling some of the available research and analytical resources in the direction of manpower research utilizing CPS data potential — can probably be best undertaken by carefully nurturing along such analytical sprouts as become discernible, and by stimulating the growth of new ones. In both instances it would seem that the availability of empirical raw material could play a significant and vital role. In particular, CPS information is here of interest because of the possible effect on analytical demand that might be exerted by the format in which information might become available. Format here is used in its broadest sense and includes the earlier mentioned Bureau of Labor Statistics experimentation with data storage and retrieval. The CPS programs have become a unique source of information on the various aspects of manpower. Because of the wide coverage of the sample, the length of time over which data have been collected, and the generally high technical level on which survey operations have been conducted by the Bureau of the Census, CPS manpower information promises to become an outstanding example for modern information-retrieval potentialities, including the further processing of retrieved data as part of the same system, developed by the Bureau of Labor Statistics.

There can be little doubt that the form in which statistics are being made available can have a pronounced effect on their analytical potential quite commensurate with the content of the information itself. This is true even where analysis is unsophisticated and, by and large, limited to description. Where the analysis proceeds along the lines of classical inference, the format of the data supply is of even greater relevance. If, furthermore, there is available to the analyst a flexible and powerful data-retrieval system, it becomes feasible for him not only to test but also to explore experimental approaches to the problem of generating hypotheses. Although it stands to reason that an effective information-retrieval system provokes the demand for data and is a long step forward in satisfying analytical curiosity, the resourceful and imaginative use of such an information system for explicitly heuristic[3] purposes has as yet unrealized possibilities. And although "heuristic" in the conventional sense of the term addresses itself to problem *solution*, there is no cogent reason for not considering similar approaches directed to the *position* of questions and the *formulation* of problems. Much too little is known about the difficult but intriguing "art of scientific discovery" and about the creation of scientific hypotheses.[4] Even less is known about the role that information systems containing statistical data could play in such a process. It would appear that electronic data processing — although no cure-all — might serve as a most promising vehicle also for heuristic exploits. At present, attempts in this direction, still in their infancy, are on theoretical levels linked to the discussion about inductive inference; on the applied level they are closely related to pattern-recognition techniques.[5] Such attempts, if applied to an appropriate data-storage system, could be made to stimulate speculation about, and creation of, hypotheses based on the data. In operations research the problem of the relation between search for hypotheses and availability of empirical information has already become acute.[6] Couched within a heuristic context, its close connection with data-processing technology has been recognized. In spite of first impressions that speculations directed toward experimental use of data-information systems not only for retrieval of information but also for assisting the researcher in his "act of creation" may have the flavor of science fiction, this newly emerging data-system potential with all its qualifications deserves sympathetic consideration by the forward-looking manpower analyst.[7]

---

[1] Among the items not regularly collated are those referring to income and educational attainment.

[2] More ambitious endeavors might be directed to including in such a system also state and locality data such as those likely to be collected under the anti-poverty program or the Vocational Education Act.

Thus far, the discussion has been entirely in terms of CPS. It goes without saying that an appraisal of the analytical potential of statistical data now available for manpower research would be quite lopsided if limited to a single source of statistical input, even though imposing. The resulting misjudgment and underrating of research potential would be more pronounced as the inclusion of additional major information sources would have a joint rather than an additive impact on their combined analytical possibilities. Among such major sources which must be considered in order to round out the present sketch are the relevant data resulting from the population censuses, from the Bureau of Labor Statistics' establishment-sample-based employer surveys (including the data resulting from joint programs of the Bureau of Employment Security and state agencies), from the Bureau of Labor Statistics' Survey of Scientists, Engineers, and Technicians on industrial payrolls, from the National Science Foundation's National Roster of Scientific and Technical Personnel, and from the operating statistics accruing in the Social Security Administration (including the continuous work-history sample data). Although this list is by no means exhaustive, it conveys an impression of the order of magnitude of the added analytical potential inherent in each source, and of the tremendous advantage in the joint and simultaneous use of supplementary statistical data.

This multiplicity of relevant statistical sources also points to another open question which has not been considered here, namely: How should considerations of analytical potential enter future specifications of basic data to be introduced? Here the judicious and ingenious development of ideas similar to those underlying the experimental efforts alluded to in connection with manpower and employment statistics, while considering not merely one particular survey program but all major data sources simultaneously, opens entirely new vistas.

A more nearly complete evaluation of analytical potential ought, therefore, to include also explorations along this just-mentioned line of statistical information programming. This would make it possible to hope for locating points where relatively small additional data inputs would generate large analytical output potential.

---

[3] G. Polya, in his stimulating classic, *How To Solve It* (Princeton, Princeton University Press, 1945), describes heuristic reasoning as "reasoning not regarded final and strict but as provisional and plausible only, whose purpose it is to discover the solution of the present problem." Lately, an entire literature has been developing on this subject. See, e.g., David B. Hertz and Roger P. Eddison, editors, *Progress in Operations Research,* New York, Wiley, 1964, Vol. 2.

[4] That this problem, which has been hanging in the air for some time, is about to come down to earth is illustrated by the publication recently of an impressive attempt to popularize the issue. See Arthur Koestler, *The Act of Creation,* New York, Macmillan, 1964.

[5] See, e.g., L. Uhr and C. Vossler, "A Pattern Recognition Program That Generates, Evaluates, and Adjusts Its Own Operators," *Annals of the New York Academy of Science,* 50, 1961, and literature listed *in* Edward A. Feigenbaum and Julian Feldman, editors, *Computers and Thought,* New York, McGraw-Hill, 1963.

[6] See, e.g., Van Court Hare, Jr., *in* David B. Hertz and Roger P. Eddison, editors, *Progress in Operations Research,* New York, Wiley, 1964, Vol. 2, p. 147: ". . . good hypotheses for trial are chosen using as criteria for 'goodness' certain available information that may suggest which possible solutions should be tried first, or that may provide quick tests for distinguishing likely solutions from unlikely ones."

[7] Rewarding suggestions are being contributed by the recently developing discussion about "artificial intelligence," going back to the early work of A. M. Turing, the famous British mathematician and designer of high-speed electronic digital computing equipment (see, e.g., his not-too-technical article, "Computing Machinery and Intelligence," *Mind,* Vol. 59, 1959, pp. 433 ff.).

Returning again to CPS, it appears in retrospect that recent trends in the new information technology seem to suggest that the discrepancy which exists now between CPS manpower data supply and analytical demand for the data could be considerably reduced by:

1. Producing information for which there is now unsatisfied demand, not only by undertaking new inquiries but by better utilizing existing data with the aid of such devices as matching, collating, and linking of information, and reducing the non-sampling error (and, if feasible, the sampling error).
2. Expanding analytical horizons through heuristic experimentation with such information as that stored on Bureau of Census tapes, utilizing trial and error methods, computer simulation, pattern search, and similar devices to stimulate analytical appetites. This process, in spite of its strongly empirical and experimental character, may point to a heavier emphasis on what should be considered basic as contrasted with applied research.

Beyond such full exploitation of existing information and facilities, it is to be expected, of course, that there will remain a demand for information neither available nor producible by skillful manipulation of collected data. Rather than satisfying such demand in a more or less haphazard way, on an *ad hoc* basis, as it happens to emerge here or there, the development of a manpower and employment data-storage and -retrieval system could become a most helpful instrument in the identification of information gaps and in the rational programming for the inclusion of new, and the omission of obsolete, information categories within the frame of the present survey system.

The aim of the apprehensive manpower analyst, therefore, should be to utilize better the existing information and to seek its improvement rather than to pin his hope and faith in the main on additions to the present store of raw materials; and to avail himself of the new information technologies for use not only in finding answers to analytical problems but also in experimenting with a view to discovering new approaches to the complex and delicate art of creating scientific patterns of thought, asking questions, and generating hypotheses.

Appendix I.

Administrative and Policy Considerations

The national resources character of government statistics and the great stake that many agencies and departments have in statistics create policy questions and problems which are directly pertinent to the question at hand, namely, the analytical potential of the Current Population Survey (CPS) type of manpower statistics.

There is the question of the preservation of the statistical information assembled by the government. It involves decisions not only on how long information is to be preserved but also on what kind of information is to be selected for preservation.

The problem is complex: the typical survey goes through several phases or production stages, involving recording of primary and of secondary or derived statistics. The first phase obviously is the recording of the basic data, of the statistical raw material, so to speak. This is usually done through questionnaire, schedule, or similar source document. Frequently more than one such record form is used since, in addition to the primary, substantive information on the respondent, there may be subsidiary survey forms and schedules for technical and administrative rather than substantive reasons. From the analyst's point of view, both kinds of records may be of interest. For example, in the CPS a control card is used in addition to the substantive schedule, the so-called FOSDIC[1] schedule. As will be mentioned later, both forms contain information that could have analytical implications.

In most surveys, data must be processed through several phases, but at least through one in addition to the phase of collecting and recording the primary data. This second phase, which is directed towards summarizing and condensing the primary data, may involve straightforward sorts and counts, subtotaling, and the like; it may involve the computation of simple quantities such as percentages; but it also may demand more complex manipulations as in the processing of data derived from sample surveys where estimates of so-called population parameters must be prepared for various population cells and various levels of aggregation; it may involve the computation of measures of sampling errors for the several classes of parameter estimates, and this in turn involves the computation of sums of squares of differences, and the like.

CPS is a sample survey — one of the best but not one of the simplest. It is based upon a multistage design involving in turn several levels of estimation.[2] Processing the results of surveys of such an order of complexity naturally leads, in addition to the record of primary statistical information, to a host of derived and higher order worksheets and summaries condensing the statistical raw material into an often impressive number of secondary figures.

A policy regarding preservation, therefore, must be based on considerations involving a great many sets of records. Each of these sets may have its own analytical potential, and whether it has been preserved may become a crucial point in question for the analyst, especially where historical records such as those in time series analysis are at issue. Closely related to the decision on whether to preserve information, and if so for how long, is the question of record storage. With the introduction of magnetic tape as a medium for recording information, the problem of storing information has perhaps become somewhat less acute because of the lesser cost of storing tape. On the other hand, recording information on tape, or for that matter on other media for use with electronic data-processing equipment, tends to widen the gap between original and derived information. If survey schedule answers were, for one reason or another, not punched in their original form, such information is likely to be irretrievable if schedules are destroyed. Processing of CPS information is now a tape-based operation, but a great number of worksheets are involved in condensing further the tape-derived material for special and exploratory purposes. Due to the division of labor on the CPS, since 1959, between the Census Bureau and the Bureau of Labor Statistics, there are actually two different sets of manpower records in existence, although the Census tape is the main source of all materials. It is now anticipated that these Census tapes will be preserved for some time to come, but not the original schedules and not the considerable variety of worksheets. Concerning already processed information, its availability over various time periods will be discussed later. Suffice it here to point out that any attempts to subject to analytical use information which is now available only on the original record would be very expensive and cumbersome, to say the least.

Among other policy problems of great concern to the analytical data-user are agency attitude and practice with respect to the question of making information available. Especially to the data-user outside of government, this question may be of critical importance. Particular questions usually do not arise primarly because of the high cost involved in producing the desired information or because of the difficulty of scheduling data-retrieval operations promptly. Rather there are here involved certain questions of principle that may preclude the analytical use of information, even though there may be little doubt about its analytical potential.[3]

---

[1] Film Optical Sensing Device for Input to Computer.

[2] For a more detailed description of the sampling design, see Appendix II.

Foremost among those problems is the so-called disclosure question. In many government surveys, answers are requested from the intended respondent under mandatory powers; in many other cases, although answers of the respondent are voluntary, he is promised confidentiality. Answers to CPS questions are of this nature and, therefore, become privileged information.

Although at first glance it might seem that for a survey such as CPS disclosure of ,respondents would be of no consequence if analysis of the data were the issue, this is not so. First of all, any inspection of records from which a respondent's identification can be inferred involves a disclosure problem. This might include use of original schedules to obtain information that has not been quoted, edited, or transferred to some other medium on which identification is impossible or from which it can easily be deleted. Then there are the cases which frequently might be of definite concern to the analyst, where it would be desirable to follow up on a subsample of respondents with questionnaire or interview to supplement the original information, either in depth and content, or in time. Related to this situation is another which is gradually becoming of considerable interest and also has significant analytical potential: the matching of records from several sources in order to take advantage of the different contents and time dimensions of, say, different survey programs and thus pool information.

There is at this time no uniform official government-wide policy with respect to disclosure of statistical data. However, several government agencies for which the problem of information disclosure has arisen in one form or another have addressed themselves to that question. The objective is, in general, to protect information received in confidence; the phraseologies used refer to the need of avoiding disclosure lest it give some persons undue advantage or disadvantage (Federal Trade Commission), and to the assurance that the results of such disclosure would not be incompatible with the public interest (Food and Drug Administration), and so forth.

A more explicit policy with respect to the disclosure of statistical information of a nature similar to the one here discussed has been developed by the Department of Health, Education, and Welfare. Since the use of Social Security records is restricted by the very enabling language of the agency (Section 1106 of the Social Security Act), it became necessary quite early in the agency's history to spell out the conditions under which information might be disclosed. In fact, it was the first regulation issued by the new Social Security Administration that concerns itself with the disclosure of official records and information.[4]

Of most direct relevance to the question here discussed is the policy with respect to Census reports. The basic rule of law[5] which enunciates this policy forbids: (1) the use of the information for other than statistical purposes; (2) the publication of the data in a way that would make it possible to identify the individual or establishment which furnished the information; and (3) the access to records by anyone other than the sworn officers and employees of the department.

A Presidential proclamation in 1929 for the 15th Census affirms, in guaranteeing confidentiality of Census reports, that "no person can be harmed in any way by furnishing the information required. The Census has nothing to do with taxation, with military or jury service, with the compulsion of school attendance, with the regulation of immigration, or with the enforcement of any national, state, or local law or ordinance. . . ." Thirty years later, another President signed into law a bill reaffirming this principle — this time because of a case which might have resulted in disclosure of information to the Federal Trade Commission. We are mentioning the various policy statements and implementations in such detail because of the considerable importance of the question for research.

These policies do exert a definite and restrictive influence on the analysis-oriented demand for information even though, as can be seen, research was not their primary objective. In fact, disclosure problems and procedures can raise issues that are, in the daily work of the researcher, quite as serious and discouraging as the nonexistence of desired information; and they can raise hurdles in the way to full exploitation of the analytical potential of existing information which are commensurate with, for instance, serious conceptual or statistical deficiencies of the data.[6]

---

[3] See, however, footnote 6, this appendix.

[4] Department of Health, Education and Welfare, Social Security Administration, Regulation 1 as amended, Part 401, Chapter III, Title 20, Code of Federal Regulations.

Except when authorized by the person who furnishes the information, the main instances in which data can be disclosed are for social security and related purposes; in situations where the national security is involved; where the office of the Attorney General requests under the Immigration and Naturalization Act information concerning aliens; and where the Treasury Department requests information which, as a rule, it had originally furnished. The only provision which, by implication, is directly relevant to research, is Section 401.3k of the regulation, according to which "statistical data . . . not relating to any paricular person . . . may be disclosed when efficient administration permits."

The details of the "manner of disclosure" of this information are left to the commissioner of the agency (Section 401.3j), and attempts have been underway for some time to develop specific guidelines. The question has again become acute more recently with the renewed efforts of the agency to promote research. (See *A Research Program for the Social Security Administration: Report of Advisory Group to the Commissioner of Social Security*, 1960, and *Report of Advisory Group Under Social Security Administration Research Program, Supplement*, 1960.)

[5] USC Title 13, Section 9.

The remedy to this situation can be found in the area of research organizations rather than in the area of research techniques. In addition to defending the justifiable data needs and share in this national resource on the part of the social sciences, while preserving the rights of the individual and the good will of the respondent, some further explorations with new research vehicles may be appropriate. For instance, avenues might be found to a closer working relationship between data-gatherer and information-storer on the one hand, and data-analyst and information-consumer on the other. Here the potential role of the middleman in the form of a research organization might well prove useful in the quest for new research vehicles — new in the social sciences — in the form of rejuvenated Project-Rand-type of arrangements between government agency and non-government researchers and others. For the same reason, attention is drawn later in this report to some recent developments in data storage and retrieval which, ultimately, may simplify some of the problems inherent in data disclosure.

Once the information has been stored, there emerge a series of problems of a predominantly administrative character in connection with the retrieval of the information. Some of the resulting practices and policies may have a noticeable impact on the analytical potential, and they must be faced, as must be the content and quality problem of the statistical information.

The Current Population Survey is based on a probability sample. Its results, therefore, are subject to certain limitations, in addition to the usual ones resulting from concepts underlying a survey,— limitations in quality of the obtained information, and the like. It is pointed out elsewhere that censorship of data is no safeguard against their misuse unless there is complete blackout of information. In most instances, government agencies seem to be sympathetic to this point of view, and seem to be reconciled that they are unable to preclude the misuse of some of the information. One might even argue that it would be dangerous, and quite impracticable, for statistical agencies to assume, even by implication, responsibility for results obtained by data-users not in their purview. Also, in the utilization of data by the analyst and researcher, intentional misuse may be a less frequent danger than the exploitation of information for other reasons. Nevertheless, CPS-type data do contain raw material that could be misused in a manner leading to highly explosive consequences. The best rule, it would seem, is a simple rule of thumb: to make available and facilitate the use of as much information on sampling and nonsampling errors of the data as possible, but to leave the ultimate responsibility for analytical results with the analyst. This is the perspective from which the problem of data quality is later viewed.

Finally, an agency's policy and practice with respect to divulging data for analytical use must be such as to safeguard efficient operation of the agency and to avoid interference — never mind how well intended. Hence a major problem which is about to emerge in connection with analytical demand for data is how to enable the data-requester to communicate efficiently with the data-supply and data-storage operation. Where already published information can serve as analytical raw material, this problem does not arise. However, when information is being requested which is not available in published form or not conveniently so, but where it must be retrieved from tape storage or otherwise, there is a very real problem of communication. This problem is later touched upon by sketching the outline of communications procedures with a view to using the Bureau of Labor Statistics' manpower statistics storage system.

The just-mentioned problems and difficulties are particularly acute for research which is to be undertaken outside the confines of the data-producing agency. Although research should not be undertaken solely outside the agency, neither should it become an agency monopoly. The need for taking advantage of the best research talent in the most efficient way would point toward the great merit of policies which stimulate research at large. Some of the earlier mentioned administrative issues and difficulties could be easily solved by such policies. Where this appears unlikely or impossible, new research vehicles may have to be explored and devised — vehicles in the area extending between agency-conducted, in-house research on one side, and the private, individual researcher on the other.

The importance of the problem of how to husband wisely statistical information and, at the same time, to stimulate its use by researchers has found expression, among others, in organizational arrangements and in the preparation of reports and the like, both within and outside the government. Their very existence and scope are of interest when evaluating the exploitation of the analytical potential of CPS statistics.

Among the organizational efforts and expressions here of importance is, first of all, the work of the American Statistical Association's Advisory Committee on Statistical Policy. This committee was established in 1951 at the request of the Office of Statistical Standards in the Bureau of the Budget, to advise the office on broad matters of public policy in the statistical area.[7] This committee is singled out because it clearly recognizes the importance to researchers of statistical raw materials produced by the government, and because the degree of availability of this national resource is likely to influence the direction and level of accomplishment of research for many years to come.

---

[6] Attitudes of different agencies vary all the way from negative and opposing to positive and constructive. For an example of the latter point of view, see Morris H. Hansen, *Some New Developments in Increasing the Availability of Census Information*, paper presented at the meeting of the American Statistical Association and American Economic Association in Chicago, Illinois, on December 30, 1964.

In this committee's report, *Availability of Federal Statistical Materials to Nongovernmental Research Workers: A Statement of Principles*,[8] it is stated at the very outset that the records resulting from federal data-collecting programs often produce statistical information "capable of statistical analysis beyond that which can or should be carried out by the collecting agency or any other agency of the Federal Government." The statement further stresses the importance of establishing the policy of allowing and encouraging the further analysis of these materials in order to obtain optimum benefit from the federal data-collecting activities. It is also pointed out that "while a priority should be given to bona fide research uses in the general public interest, special tabulations should be permissible for all legitimate uses." For research projects in the general public interest, it is suggested that research needs might be met by allowing even nongovernmental workers to work with the raw materials, worksheets, etc.

Two other committees address themselves to the specific problems of data storage and retrieval. One is the American Statistical Association's Committee on Data Sources and Information Systems — a committee resulting from the merger, about five years ago, of two separate committees; the other is the Social Science Research Council's Committee on Preservation and Use of Economic Data — a committee appointed about three years ago in response to the American Economic Association's consideration, in 1959, of the problem of the preservation and use of data for economic research. Whereas the American Statistical Association committee addresses itself to the broader problem of statistical programming under the impact of the "information revolution," the Social Science Research Council committee focuses more narrowly on the problem of preservation of historical data by storing government statistics on tape for use by researchers. The problem of storage, simple as it may seem, immediately raises two subquestions: Should only published summary data be stored on tape? (This would permit considerably easier, more flexible, and more ambitious analysis of the data than if only the published tables were available to the analyst. On the other hand, this format of recording information, the macrotape, although much more convenient, would not add anything to the store of information that had been previously available to the analyst.) Or should the magnetic tapes contain all information on the level of schedule detail? (The preparation and storage of such microtapes would be a much more complex undertaking, requiring possibly a good deal of additional editorial work, and raising some rather intricate problems where the basic information was derived from schedules resulting from a sample survey;[9] here the raw material before being exposed to analytical use would have to be subjected to appropriate estimation procedures.)

In essence, both committees are concerned with facilitating the prospective analyst's use of data collected by government agencies. However, rather than exploring problems inherent in the data, the committees are emphasizing information-storage and -retrieval aspects; in other words, they are concerned with the problem of how best to get the right information to the researcher. This, of course, involves a host of subproblems such as the one mentioned above in connection with sample surveys.

In addition to these general and broad-gauged attacks on the problems of analytical use of government statistics, there are the more narrowly conceived attempts aimed at manpower statistics in particular. Beginning with the 1950's, much valuable work has been done in this direction by several committees, commissions, and other groups and organizations — and of course also by the collecting agencies.

In the summer of 1954, a special advisory committee on employment statistics that was appointed by the director of the Census Bureau on the occasion of introducing a revised CPS sample, which expanded the earlier sample of 68 primary sampling units to 230, submitted its report,[10] the purpose of which was to inquire into certain discrepancies between the old and the new series. In preparing its report, the committee had to delve into related issues which raised questions concerning concepts, sampling design, estimation procedure, sampling and nonsampling errors, and the like; the report, therefore, is one of the most comprehensive statements pertaining to the earlier CPS information. After a laudatory appraisal of CPS, a recommendation is made on the last page of the report that, in view of the complexity of the problem and its great importance, a continuing committee or commission be established to inquire into the various aspects of CPS.

Because of the political sensitivity of the Current Population Survey's labor force estimates, several government committees were, at one time or another, involved in appraising CPS statistics.[11] The documents produced by these groups are among the most valuable and interesting information on CPS, illustrating the excellence as well as the limitations of the manpower type of information derived from that survey.

---

[7] *Statistical Reporter*, May 1957, p. 79.

[8] Adopted by the committee, and transmitted to the Office of Statistical Standards early in 1957.

[9] Concerning CPS tapes, it should be noted that information needed to weight the data is contained on the tapes. The only "weighting" problem that remains, if it is considered important in a particular situation, is the adjustments resulting from the composite estimation procedure which require "weighting" involving pairs of microtapes for adjacent months (see page 8).

[10] "The Measurement of Employment and Unemployment by the Bureau of the Census in Its Current Population Survey" (also known as the Stephan Committee Report), 1954.

Most recently, at the end of 1962, the well-known and highly regarded Gordon Committee report was published;[12] it probably represents the most complete and comprehensive appraisal of any single statistical program of a federal government agency. It addresses itself to definitions and concepts, to sampling design, to interviewing and related questions, to statistical adjustments of the data, and to the presentation of the results. The appendix material contains detailed descriptions of the results of the 1955 experimental study on unemployment,[13] of the sampling procedure employed in the Current Population Survey,[14] of the 1961-1963 revision of the CPS sample, and of comparisons and matching studies between CPS and population census. Thus, this report is a unique and invaluable source of information on the Current Population Survey data-producing effort.

Looking over the impressive array of documents throughout the last decade or so, one inevitably realizes that probably more work has been done in connection with the appraisal of the Current Population Survey as a source of labor force information than for any other single survey program. In spite of this, the actual analytical use of the data, its intensity, level of sophistication, and breadth must, to many, appear rather disappointing.

In conclusion, it would seem that two major lines of attack have been developing. One is illustrated by the first group of committees, formed by the American Statistical Association and the Social Science Research Council, which directed their efforts toward readying government-collected statistical information for more general analytical exploitation. They have been addressing themselves to problems of data storage and retrieval in their most general form; they are prepared to consider some of the technical as well as organizational and administrative aspects; and they recognize the crucial role that access to good empirical data plays in the research laboratory of the social scientist. More fundamentally, they realize the acute need of promoting research within a climate favorable to having government-collected data made more generally available and more widely utilized, as expressed in the earlier quoted report by the American Statistical Association's Committee on Statistical Policy. However, these committees are not actually concerned with manpower statistics as such.

The other line of attack is represented by the second group of committees; they emphasize directly the manpower aspect of the information, but their primary concern is with the evaluation and the improvement of manpower statistics collected by the government rather than with the promotion and facilitation of analysis and research based on this statistical information. What is needed in order to promote successfully the analytical exploitation of CPS data is a combination of both points of view into a single line of attack.

Thus, it becomes obvious that analytical potential of statistical data is contingent not only on the existence of pertinent data, in desired detail and of appropriate quality, but also on the availability of the information to the analyst. The analyst faces not only obstacles in the form of limitations in content and quality of the data but frequently also administrative and policy hurdles which must first be overcome before fruitful analysis can be undertaken. In view of the nature of CPS data, the kinds of problems for which solutions must be found prior to realizing the full analytical potential of the data result primarily from government policies with respect to the disclosure of privileged information, from destruction and disposal of statistical material, and from the need for providing the researcher with statistical information which may not be available in published form at all. These problems, although not statistical in nature and not predicated on analytical problems and issues, must nevertheless be kept in mind because of their direct effect on the actual realization of the analytical potential of government statistics in general, and of CPS information in particular.

---

[11] Among them were the Subcommittee on Economic Statistics to the Joint Economic Committee (formerly the Committee on the Economic Report); the Special Committee on Unemployment Problems, which was established in 1959 under the chairmanship of Eugene J. McCarthy, and which upon completion of its report in March 1960 recommended the establishment of the Subcommittee on Employment and Manpower of the Committee on Labor and Public Welfare, also known as the Clark Committee.

[12] U.S. President's Committee To Appraise Employment and Unemployment Statistics, *Measuring Employment and Unemployment.*

[13] Subsequently, findings of the CPS Reinterview Program, 1956-1961, were published. See U.S. Bureau of the Census, *The Current Population Survey Reinterview Program: Some Notes and Discussion,* Technical Paper No. 6, 1963.

[14] See also U.S. Bureau of the Census, *The Current Population Survey: A Report on Methodology,* Technical Paper No. 7, 1963.

Appendix II.

Qualification of the Data

Current Population Survey (CPS) data, being derived from a probability sample, result in estimates for which the level of precision can be estimated from the sample itself. The level of precision of the estimates, expressed in terms of sampling errors for the various estimates, is a function not only of the sample size and sampling plan but also of the estimating procedure. Although these sampling errors vary from survey to survey, that is, from month to month, the variations have not been disturbing as far as CPS is concerned; therefore, for most substantive analysis of the data — substantive as opposed to methodological — only the major changes in sampling error need usually be considered. Such changes occur when size of sample, sample design, and estimation procedure are markedly modified.

For estimates derived from CPS, measures of estimated sampling error have been published since the inception of the survey in one form or another. In view of the many different variables and attributes involved in this survey, the publication and, because of its cost, also the computation of measures of the sampling error have proceeded along selective lines; with the progress of time, a good deal of study and experimentation have led to a substantial expansion and deepening of the knowledge and understanding about variability of CPS quantities and estimates. Since CPS is based on a probability design, the chances can be computed that a sample estimate will differ by a given amount from the true value which might have been obtained if a census had been conducted.

Where forecasts must be made for policy purposes on the basis of current figures in order, say, to identify a turning point in the unemployment situation, the penalty of a wrong decision can be heavy, and the decisionmaker may not wish to assume a risk of error as high as, say, one out of three. Conversely, where analysis is based on many successive observations in an attempt to reconstruct past behavior and on search for relationships and other patterns suggested by the data, but where the analyst's intuition is also affected by considerations based on other information and on previous analytical experiences, the risk of making a serious error by using only one set of observations — those derived from the particular sample — would seem to be much smaller. In such a situation, it is neither simple nor straightforward to arrive at a quantification of the gravity of an error in order to form an opinion about the permissible level of risk, and therefore about the "confidence interval,"[1] in applying the estimates to the particular analysis. Neither is it easy to choose between the risk of avoiding a conclusion that may be correct, as well as of great analytical import, and the risk, on the other hand, of arriving at a conclusion that may later turn out to be wrong. These are decisions which only the analyst can make, and for which he and only he must assume the responsibility.

For instance, a simple-minded answer to the question of whether cells resulting from cross-classifications are or are not "sufficiently" large can actually be harmful; and it is most likely to be less helpful than sensitive awareness of the nature and order of magnitude of sampling variabilities and of their effect on the interpretation of patterns suggested by sample data. If imaginative analysis of manpower data is to be achieved rather than simple adherence to the mechanical application of numerical rules of thumb in the interpretation of sampling error, the researcher will have to be aware of the many and complex effects of the variability of the estimates from sample to sample; hence, he must use the information concerning the estimated size of the variability as well as the other many bits and pieces of experience with a great deal of resourcefulness, care, and courage. This need not guide the more nearly action-oriented interpreter of the information who, first of all, must avoid erroneous decisions, frequently engendering policy formulation and practical action based, as a rule, on the latest information. The difference between the two data uses is somewhat reminiscent of that between the researcher in pharmacology and the person prescribing drugs to patients. In more nearly statistical terminology, the risks involved in the decisions and their payoff, their cost, and their value differ from each other widely in the different situations.

In order for the analyst to form an opinion about the magnitude, hence the influence on his conclusion of sampling variability in CPS data, a nomogram is here given

---

[1] "Confidence interval" is the name given to a range about the sample value within which the true value is believed to be found with a stated probability, say, p. In other words, the statement that the population value lies within a particular confidence interval will be true in a proportion p of all possible cases. Hence, the wider the confidence interval, the higher the p, all other things being equal; but the wider the confidence interval, the less useful the information derived from the sample, in that it must be expected that the population value which is to be estimated from the sample may differ considerably from the sample estimate — and it is only this (unknown) population value that is of real interest to the analyst, not its sample estimate as such.

(see figure 1, page 52) which summarizes the relationship between sample size and sampling error as applied to CPS. In addition to size, sampling errors also depend on the sampling design, including the estimation procedure, and, of course, on the variability of the particular magnitude in the population as a whole. Hence, strictly speaking, there is a sampling error for each observed variable and attribute. Also, sampling errors vary from survey to survey — in this instance from month to month. However, it appears that for most estimates of Current Population Survey variables their patterns of sampling variability are similar;[2] the same is, broadly speaking, true for these patterns over various points in time; but as is to be expected, this is not so for estimates derived from Current Population Surveys which differ from each other in the nature of their sampling design or in other major respects. Each major modification in such aspects has, as a rule, brought with it a different sampling error pattern. As a matter of fact, most of the changes and modifications in the CPS have been aimed directly and quite explicitly at a reduction, hence a change of the sampling errors.

Figure 1 describes in the form of a nomogram the change of the sampling error with changing sample size. In particular, it shows how the sampling error increases as the size of the evidence decreases, that is, as the number of sampled persons to whom a given estimate is referring becomes smaller. It must be noted that these measures of sampling variability refer to estimates of levels as against estimates of changes from survey to survey, that is, from month to month. The sampling errors of these intermonthly changes are smaller than for the levels. (See footnote 4, this appendix.)

For the sake of the analyst who may not recall some of the intricacies of statistical method, it should be pointed out that the measure of sampling error used in figure 1 is a relative measure — the ratio of the so-called standard error (the standard deviation of the sampling distribution of the estimate) to the estimate itself. For estimates of totals, for instance, the sampling error would be compared with, related to, or divided by the estimate of the total. The practical advantage to the analyst of such a measure results from the fact that this measure is in the nature of a pure number which can be interpreted independently of the quantitative character of the variable in question and of the units in which it happens to be measured. Statistical convention gives this relative measure which is comparable to the sampling variance the name "relvariance," and to its square root which is comparable to the standard error, the name "coefficient of variation."

This coefficient of variation can be interpreted as being analogous to the standard deviation of the sampling distribution. It is usually given in integral percents. Thus a coefficient of variation of .5 means a coefficient of 0.5 percent, and one of 12 stands for a coefficient of 12 percent. On the assumption of probability theory, the population value that would have been obtained in a complete enumeration of a census type is to be found within a range about the estimate derived from a probability sample; this range extends from plus to minus the percent indicated by the coefficient of variations or a multiple thereof. For example, if we find the coefficient variation equal to 12, this statement would suggest that we look for the "true" value of the characteristic somewhere in the interval extending from 12 percent below to 12 percent above the value of the sample estimate in the particular instance.

A statement such as the one just made will be valid in the long run a little more often than two out of three times; if it is felt that the risk implied by these odds is too high, and if consequently the analyst feels that a risk of being wrong only one out of 20 times is all that he is willing to assume, the value found for a particular coefficient of variation would have to be multiplied by two — that is, doubled. If the coefficient of variation were tripled, the odds that the resulting statement as to the range within which the population value is to be found is valid would be about 998 out of 1,000. In other words, subject to the assumption of classical probability theory, the coefficient of variation can be interpreted in an analogous way to the standard deviation in a normal distribution.

In order to find the approximate numerical value of a given coefficient of variation from the nomogram in figure 1, the analyst must first decide on the appropriate size of sample cell to be used. If the analysis proceeds on highest levels of aggregation, and if the estimates refer to the entire sample, then the size of the entire CPS sample is obviously the appropriate size to use.

Next, the analyst will find the intersection of the vertical line corresponding to the desired sample size (as indicated on the horizontal axis of the nomogram) with the sloping line representing the particular Current Population Survey period to which the estimate refers.

Finally, the analyst will look up on the right margin of the nomogram for the value corresponding to the just-located intersection. This value, if added to and subtracted from the sample estimate, gives the percentage range within which to find the true value. In so doing, the analyst, as already mentioned, will be right in the long run in two out of three cases or somewhat more often.[3] As can be seen, the nomogram consists of four

---

[2] See Joseph Waksberg and Robert B. Pearl, ***The Current Population Survey,*** a paper given at the annual meeting of the American Statistical Association, Chicago, 1964.

[3] For instance, if the analyst is interested in a particular labor force category which according to a current survey turns out to be in the neighborhood of 500,000, he will trace a perpendicular above the number 500 (thousand) on the horizontal axis; and from the point where it intersects the

(Continued on page 51)

sloping lines, each of them corresponding to a specific period in the history of the Current Population Survey. The lower the curve, the smaller the coefficient of variation for given sample sizes. The highest curve therefore describes the situation with respect to sampling variability during the early years of the Current Population Survey, immediately following the conversion, late in 1943, of the sample to a complete probability design. The lowest curve describes the situation since the latest major revision of the sampling design, that is, from April 1962 to date. Although these curves cannot represent exact regressions of coefficient of variation on sample size, they are of assistance to the analyst who wishes to form an opinion as to the order of magnitude of sampling error as well as to its general pattern.

The analyst who is concerned with the interpretation of current or relatively recent information can find his way through the maze of sampling variances by devices such as the nomogram shown in figure 1. Similarly, if he is interested in a past situation, he will secure the estimates of sampling variances as they obtained at that time.[4] Somewhat different must be the considerations that enter if several successive surveys are involved which go through a period of several modifications of the sample. Generally speaking, the history of the CPS sampling errors is one of persistently declining numerical values. The fact testifies to the persistence, ingenuity, and success of those responsible for the operation of the Current Population Survey, but it is apt to bring with it one or the other complication for the analytical user of the estimates.

First of all, in forming an opinion about the order of magnitude so as to put the problem in its proper perspective, it is useful to observe this reduction in sampling error measures for the Current Population Survey from the beginning of the survey to its present state. The corresponding gain in precision will, of course, differ from variable to variable, since their sampling errors differ; therefore, they are variously affected by the several changes in sampling design, including number of primary sampling units in the sample during a particular month or period under consideration. However, since such comparisons must of necessity be undertaken on the basis of rough approximations and conjectures, excessive conceptual refinement would be inappropriate and misleading. Therefore the effect of design improvements on two or three major categories of variables is sufficient to assist the analytical consumer of the data in his delicate task of appraising the relative precision of his statistical material. Looking at three major categories of estimates — estimates of nonagricultural employment, of agricultural employment, and unemployment — it appears that for equal sample size and sampling variances of the first CPS design that was based on a proper probability sample estimates were higher than the present ones by about 80 percent for nonagricultural employment, by about 150 percent for agricultural employment, and by about 50 percent or better for unemployment.[5]

The most important single contribution to reduction in the variance is due to the increase in sample size which was introduced in 1956. This improvement alone may account for about one-half of the reduction in the sampling variances of the estimates for nonagricultural employment, over 85 percent for agricultural employment, and over 75 percent for unemployment. The remaining improvements in precision of nonagricultural employment can be accounted for mainly by the use of the composite estimation procedure since early in 1954, and by the recently introduced use of list sample segments. Each of these modifications contributed about one-eighth to the reduction in total variance for nonagricultural employment. Finally, each of the following factors helped to reduce sampling variances of the estimates by about one-tenth: the introduction of exact weighting procedures on electronic data-processing equipment, instead of the use of schedule substitution and multiplication for imputation and similar adjustment of the survey results; and the use of an improved stratification system. In importance, these latter two improvements ranked next to the increase in the number of primary areas in terms of variance reductions for agricultural employment and for unemployment estimates; their contribution to the total decrease in sampling variance accounted for about one-fourth for agricultural employment, and about one-tenth or better for unemployment. Each of the two improvements shared about equally in the resulting gain in precision. Thus the analyst who wishes to reconstruct economic time series will have to be more on the lookout for changes in sampling precision when directing his attention to agricultural employment and related series than when analyzing nonagricultural employment and unemployment. These remarks are intended to impress the prospective analyst with the need for considering sam-

---

lowest of the four sloping lines (for the CPS phase April 1962 to date), he will trace a horizontal line toward the right; such a line would meet the right-hand perpendicular scale, the one for coefficients of variations, about three-quarters of the way from 6 to 7. He, therefore, will estimate the coefficient of variation to be about 6.75 percent. This means that the particular labor force category may be anywhere between 500,000 less, say, twice 6.75 percent, and 500,000 plus twice 6.75 percent. Put differently, the analyst looks for the true value somewhere between 567,500 and 432,500, and he expects that such a statement will be correct 19 out of 20 times. For a corresponding estimate from a sample of size 50,000, the coefficient of variation found by inspection of the nomogram would be slightly over 20 percent.

[4] For more exact sampling variances, see the various publications of the data, such as the *Monthly Report of the Labor Force* and *Employment and Earnings*, or the various CPS publications by the Bureau of the Census (P series).

[5] These remarks are based on an internal Census memorandum of February 18, 1964.

Figure 1

Approximate Relvariances by Size of Estimate for Various Historical Phases of CPS

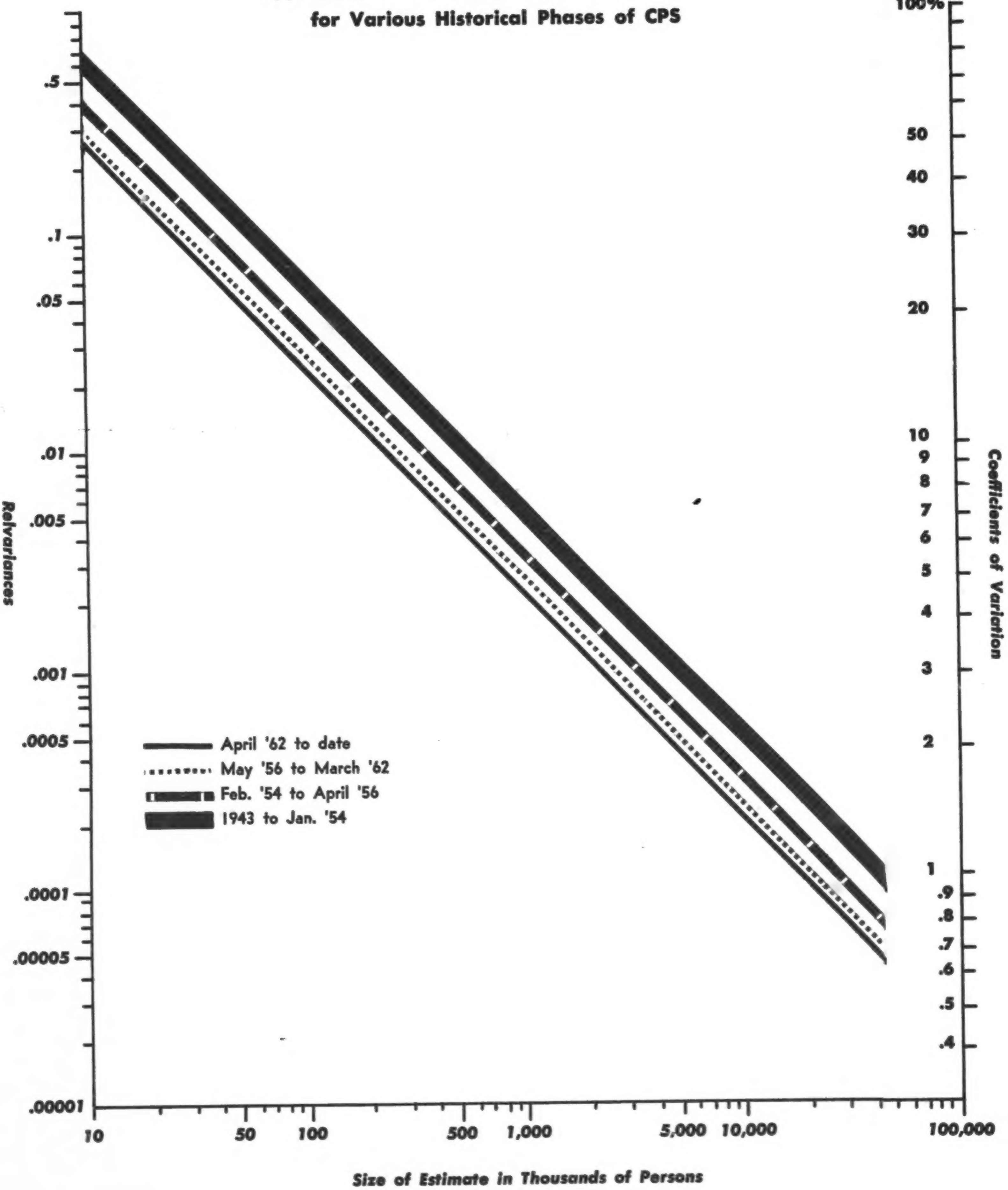
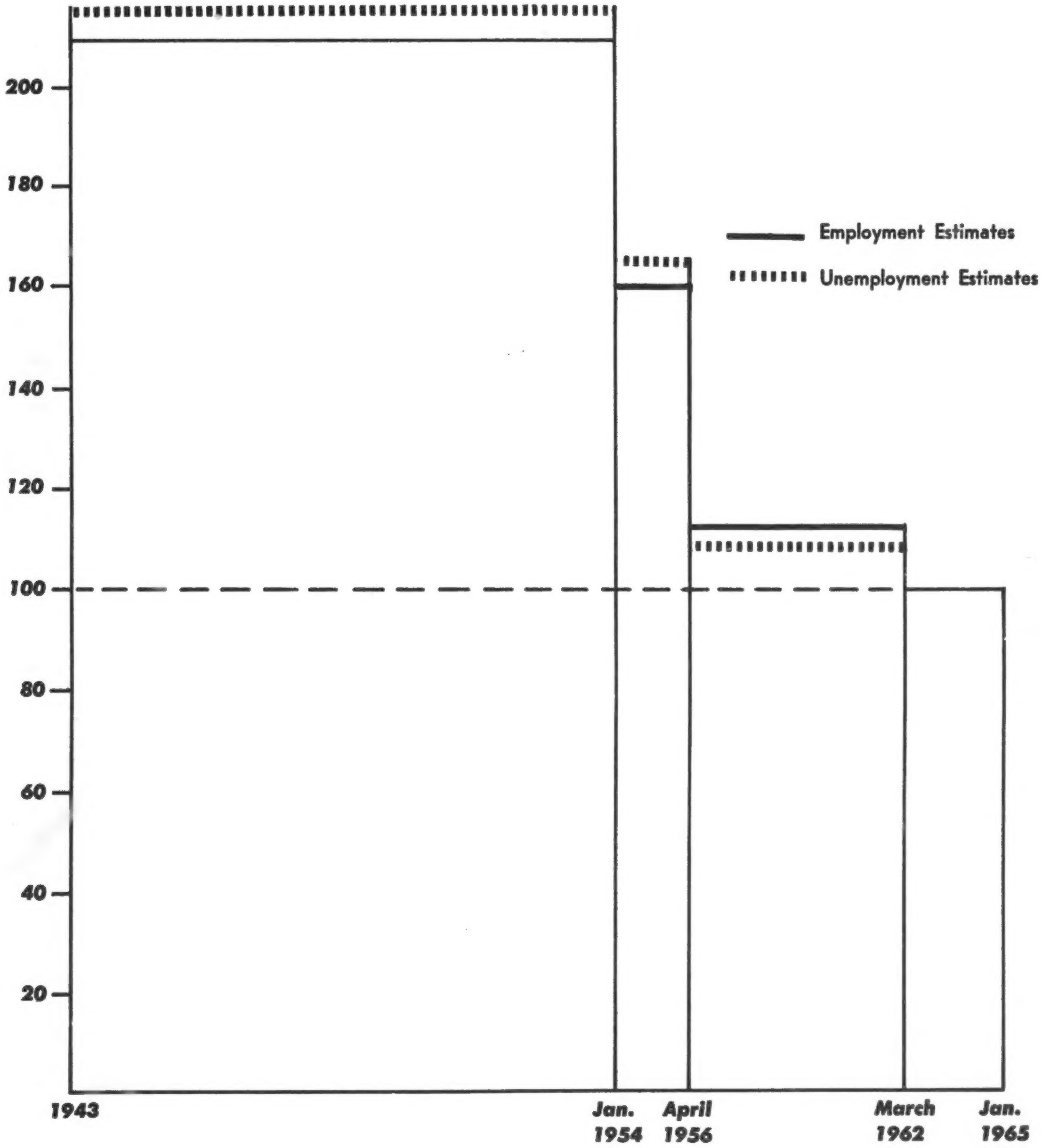


Figure 2

Successive Reduction of CPS — Relvariances

(Current Relvariance = 100)



pling error in his analysis and also with the fact that as between the various variables the differences are probably not sufficient to be of importance in the usual kind of economic analysis. This, of course, is not true when the emphasis is on time series analysis since here the differences over time in precision are considerable. In order to convey a general idea of this change over time, the diagram in figure 2 (see page 53) was prepared; in this diagram the higher a particular bar, the larger the relvariance, that is, the square of the coefficient of variation, and the lower the precision of the estimates. Relvariances are here again used in preference to simple variances because of the great shifts in the level of the variables over time. The scale on the left side is to be interpreted in the following way:

1.0 or 100 percent is the basis of the comparison and reflects the level of relvariances at present; as we go back into the past, these relvariances increase according to this scale. Thus, from 1943 to 1954 these relvariances were over two times as high as they are at present. The diagram also conveys the fact that the largest reductions in sampling error took place prior to 1956. This is only natural because it becomes gradually more difficult to reduce sampling error beyond a certain point unless the sample can be substantially expanded. Moreover, it can be seen from the diagram that the relvariances have not greatly deteriorated if we remain in the period from 1956 to date; this would seem important especially in view of the fact that, as mentioned earlier, most micro information is available only since 1959. Finally, figure 2 shows that for the two — and conceptually rather extreme — variables given, employment and unemployment, the differences in the reduction of the variance are not excessive.

So far the discussion has centered on sampling errors in connection with the regular monthly Current Population Survey. As to supplemental surveys, the standard errors for broad totals are generally about the same as the standard errors of average levels of monthly estimates for the Current Population Survey at large.[6] Similarly, standard errors for year-to-year changes of estimates are also approximately equal to those for levels of estimates. There is no analogue to the standard errors of month-to-month changes of estimates derived from the general Current Population Survey because the supplemental inquiries underlying the Special Labor Force Reports are conducted on an annual or irregular basis. If, therefore, data derived from such supplemental surveys were to be merged with labor force and manpower information derived from the CPS proper, the analyst could be guided by the same general considerations when interpreting sampling errors for supplemental surveys as when using estimates derived from the general CPS.

At this point it may perhaps be appropriate to warn against a pitfall. In considering monthly data the analyst may find that the relevant measures of sampling error are too high for his purpose. One approach, therefore, is to merge the monthly data into annual averages. It is quite obvious that if such a procedure is analytically satisfactory, it will yield smaller sampling errors than are attached to the monthly data. However, because of the rotation feature inherent in the design and also because of the substitution of neighboring households for the households leaving the sample, the gain from such averaging is considerably less than, say, 12 times. The effect on the sampling variance of such a merging will depend, among other things, on the month-to-month correlation of the variable under observation (the gain will be less, the higher this intermonth correlation), and on the intraclass correlation of the households in the selected primary sampling units for the particular variable (the lower the intraclass correlation, the higher the gain that might be achieved).[7] In general, it would be unwise to expect from such a collapsing of the sample an improvement in the relvariance by more than a factor of five.

Before leaving the topic of sampling error, a remark may be in place on the future outlook concerning sampling variances. At the end of 1962, the President's Committee To Appraise Employment and Unemployment Statistics recommended that the CPS sample be increased tenfold. Such an increase may seem overly optimistic, but it is by no means impossible or even unlikely that the Current Population Survey may be expanded in the foreseeable future.

Since the sampling errors of the estimates vary inversely as the square root of sample size, other things being equal, a quadrupling of the CPS sample would be synonymous with reducing the sampling error to one-half. In reality, the gain will most likely be greater than that since it is assumed that a substantial expansion of CPS would increase the number of primary sampling units and would introduce design improvements as well as expand the number of interviews. Also, the ability to show tabulation detail would increase not as the square root but directly as the size of the sample.

---

[6] These measures of sampling variability are published in the explanatory note section following the text of the Special Labor Force Reports. Prior to the beginning of this publication in February 1960, the material was published in CP Report Series P-50 by the Bureau of the Census.

[7] Intraclass correlation measures the correlation of elements within a particular cluster relative to the variability of the elements over the entire population. It can be used to determine the precision of estimates from a given cluster sampling design, such as CPS; if the measure of intraclass correlation (a Greek rho, P in customary notation) is larger than zero, the cluster sampling variance will be larger than the variance which would have been obtained in an unrestricted random design. An increasing P, from zero to one, indicates that the clusters are becoming more and more uniform. See, for instance, Morris H. Hansen, William N. Hurwitz, and W. G. Madow, *Sample Survey Methods and Theory*, New York, Wiley, 1953, Vol. I, p. 259; on its implications for the kind of sampling problem discussed here, see also Morris H. Hansen and William N. Hurwitz, "Relative Efficiencies of Various Sampling Units in Population Inquiries," *Journal of the American Statistical Association*, Vol. 37, 1942, pp. 89 ff., and W. Edwards Deming, *Sample Design in Business Research*, New York, Wiley, 1960, pp. 122 ff., and p. 483.

Such an improvement in precision of the estimates would bring within the analyst's horizon a host of new possibilities along the line of multivariate cross-classification and analysis and opportunities to observe the manpower structure and behavior on substantially lower levels of aggregation than now feasible. The forward-looking analyst and researcher, therefore, might, as do the the entrepreneurs and progressive technicians, devise some ingenious ways to utilize and exploit the substantially improved materials when they become available.

In deriving estimates from a survey so accomplished and so highly developed as the Current Population Survey, the problem of nonsampling errors is more important to most analysts than the problem of sampling errors.[8]

No theoretical model comparable to the models for sampling errors from probability samples is available for the study, interpretation, and insurance against nonsampling errors. Furthermore, it is not at all unlikely that some of these nonsampling errors may exceed the sampling errors considerably. The entire issue is therefore of great interest and practical importance for at least two reasons:

1. The problem of nonsampling error is itself a significant and, it would seem, rewarding object for research and analysis utilizing CPS data. This point will be made more explicit later within the context of the analytical potential and use of the data.
2. The identification of nonsampling errors, their appraisal, and, where possible, some indication of their direction and order of magnitude are of great practical consequence to the analyst and interpreter of CPS information in his effort to utilize the data properly so as to avoid incorrect descriptions and invalid inferences.

Nonsampling errors are irksome not only because they are likely to introduce bias in the survey estimates, as do other forms of error; but also because, unlike sampling errors, they are generally of unknown magnitude, and their behavior pattern is much less likely to be accessible to description by models as complete and cogent as random sampling errors. Nevertheless, a good deal of work has been undertaken in this area with respect to CPS data, and considerable experience has been accumulating since nearly the beginning of the Current Population Survey.[9]

Major and more recent systematic efforts have been undertaken and have already provided findings of invaluable help to the thoughtful data analyst, and much more is yet to come.[10] In the spring of 1963, a research program was initiated by the Census Bureau which may be described briefly as having two components: a methods-test part to test the effect of a variety of alternative enumeration techniques on labor force measurements, and a second part addressing itself to various approaches to the noninterview problem. In addition, work is continuing on the reinterview program, and Census-CPS matching projects will likely continue into the future. All these efforts are already producing a wealth of materials and cues, shedding light on nonsampling errors in the Current Population Survey.

With this steadily expanding flow of information in view — a development which is of considerable consequence for the analytical user of CPS data — it might be helpful to identify some of the more controversial problem areas of suspected presence of nonsampling errors within the confines of the Current Population Survey. And, as the various research efforts now underway are enlarging our knowledge about nonsampling errors and their magnitude, the interested research worker will be in a position to utilize the forthcoming information for the qualification of the data whose analytical exploitation he is contemplating.

For instance, the very measure of labor force participation, which depends on the classification of an individual within or outside the labor force, has been suspected for some time. From matching studies undertaken to compare the 1950 Census with CPS results at that time, it emerged that experienced interviewers will tend toward classifying more persons as members of the labor force than will inexperienced ones.[11] The more experienced CPS interviewers found an estimated 2.4 million persons in the labor force who were not so classified by the census enumerators. Similarly, CPS interviewers had a greater tendency than census takers to classify an individual as employed; this discrepancy in classifying persons was especially pronounced with respect to teenagers and elderly persons. The resulting divergence in labor force participation rates between the two surveys was stronger for nonwhites than for whites; the discrepancy is more or less consistent for various age-specific groups and not only for overall totals. Similarly, this difference between census enumeration and CPS results was more pronounced with respect to females than to males. In the 1950 Census-CPS matching study over three million women were classified as members of the labor force who were not so classified by the census enumerators.

---

[8] At this point, attention should perhaps be drawn to the fact that the analyst, when interpreting the empirical data, should be concerned about bona fide irregularities in actual economic behavior — the erratic real-world occurrences as they were termed by Julius Shiskin, "Signals and Noise," *Fortune*, February 1964 — as well as about sampling and nonsampling errors. Such considerations are especially relevant when the data are subject to time series analysis. The Census Bureau has undertaken imaginative and stimulating research to quantify the relative effect of such erratic fluctuations. (See also Julius Shiskin, "How Accurate?" *The American Statistician*, October 1963, pp. 15 ff.)

[9] For very useful information and hints illustrating certain earlier efforts, see Gertrude Bancroft, "Current Unemployment Statistics of the Census Bureau and Some Alternatives," *in The Measurement and Behavior of Unemployment*, a conference of the Universities — National Bureau Committee for Economic Research, Special Conference Series No. 8, National Bureau of Economic Research, Princeton, Princeton University Press, 1957, pp. 63-119; see also her Census monograph published in 1958 under the title, *The American Labor Force: Its Growth and Changing Position.*

[10] See, for instance, the Census Bureau's Current Population Survey Reinterview Program as described in Appendix L of *Measuring Employment and Unemployment*, the report of the U.S. President's Committee To Appraise Employment and Unemployment Statistics, 1962; and U.S. Bureau of the Census, *The Current Population Survey Reinterview Program: Some Notes and Discussion*, Technical Paper No. 6, 1963.

Considering that even in the earlier mentioned 1950 Census-CPS matching study about 92 percent of the individuals in both surveys were classified consistently, such discrepancies as observed would hardly be disturbing where analysis of overall data is contemplated. However, if analysis is envisaged which aims at some of the above-mentioned labor force subgroups specifically, the analyst might have to form an idea about presence and approximate order of magnitude of nonsampling errors on the basis of such materials as are forthcoming from the earlier mentioned research projects.

Another category which may be suspected of being plagued by nonsampling errors is occupational classification, largely because the interviewers must often rely on answers provided by household members rather than the workers. Contrary to the role of sampling errors in determining labor force participation, earlier matching study materials suggest that here men are more frequently misclassified than women. In the 1950 Census-CPS matching study, about 90 percent of the women but only 77 percent of the men were classified consistently by the two surveys. The greatest discrepancies were observed for the professional and managerial groups, and for farm laborers.

There is, of course, an unending series of possible sources for nonsampling errors in any survey. Beginning with the data-processing aspect, which involves recording, coding, editing, and computing,[12] nonsampling errors may creep into the final results at nearly every stage. Fortunately, the Current Population Survey is one of the qualitatively outstanding survey operations in which the just-mentioned errors are practically negligible and subject to a continuing quality-control program. But there are other error sources whose influence may not as yet be fully controlled. Whether they are likely to influence the results of an analysis if not accounted for explicitly is a matter of case-to-case decision.

There are errors possible because of incomplete coverage, because interviewers miss a sampling unit, because they fail to list completely all dwelling units in a designated part of an area sample,[13] or because of other forms of noninterview.

Then there are the many cases where the interviewee finds it difficult to be accurate and consistent in applying definitions and concepts. Although a good deal of attention has been given to such problems, solutions are found only gradually; a complete and absolute solution may never be found. For instance, should a fisherman who works on shares nine months of the year and does no other work during the remaining three months, because no other work exists in his case, be classified as an unemployed member of the labor force or is he not in the labor force at all? What about the housewife who works during the relatively short time that produce is to be processed and who returns thereafter to her household chores?[14] There is no doubt that in the long run such cases can be caught and experience can come to grips with them although on a case-to-case basis only. What is important is the possibility that in the meantime they may result in differential treatment and in different decisions by different interviewers, and thus become the source of nonsampling, and noncompensatory errors. The probability of this happening is of course the greater, all other things being equal, the smaller the group or subgroup is for which estimates are to be used in the analysis.

To explore such and similar suspected sources of nonsampling errors more directly, the Census Bureau has been undertaking several efforts like the ones mentioned earlier. Among them the two major efforts are the Current Population Survey Reinterview Program initiated in 1956, and the program of research into methodological problems of the Current Population Survey which was started in the spring of 1963; although highly tentative, the results of these efforts merit careful attention by the data analyst since they suggest what kind of error is likely to be involved and since, in several instances, they also convey some idea about the order of magnitude involved. Thus, the methods-test project, in spite of its short history, has explored nonsampling

---

[11] Bancroft, *The American Labor Force: Its Growth and Changing Position,* pp. 162 ff.

[12] Bancroft, *in The Measurement and Behavior of Unemployment,* p. 73.

[13] *Loc. cit.* She quotes a rate of the order of 1.3 percent of the total population for this kind of error, and another of about .5 percent accountable for by persons missed who are living in quarters visited by the interviewer; a further 3 to 5 percent because of impossibility to contact a prospective respondent before the assigned deadline.

[14] *Loc. cit.*

7

error situations in several instances.[15] Some of the suspected sources of error are:

1. The Current Population Survey seems to overstate the number of persons who indicated that they were working a normal working day, that is, 35-40 hours per week.[16] The reported average hours of work are affected in one direction by respondents who underreport time taken off, and in the other by those who underreport the hours of overtime. On the basis of analysis of the results, the methods-test tentatively suggests that the regular CPS interviews seem to result in more persons being in the category "working 35-40 hours" by probably one-tenth of the total of persons at work, that is, perhaps a little over one-fourth more persons than are actually in this category.
2. Another suspected source of nonsampling errors is the answers to the question: "How many weeks has .... been looking for work?"[17] The answers determine the estimates of length of unemployment. Here the findings were less conclusive, and it is expected that probing will continue.
3. A third methods-test concerned itself with the question whether the estimated number of self-employed is perhaps noticeably overstated because owners of incorporated business often report themselves as self-employed.[18] In other words, the CPS concept of self-employment, which apparently is also held by most researchers, is not shared by the respondents at large. The results of the test procedure do suggest the presence of such a bias, but further probing and testing will be necessary in order to arrive at an estimate of the likely order of magnitude of this nonsampling error.
4. Another part of the methods-test is concerned with what in fact is the CPS concept of labor force. According to that concept, persons are included as members of the labor force who worked anytime during the survey week for wage and salary or as self-employed persons. Such persons or, for that matter, persons looking for work, would be included in the labor force even if the work in question were casual and at very low pay. It is suspected that in actual survey practice such persons may be overlooked as members of the labor force because they may not consider their work as constituting employment owing to its intermittent nature and its low pay.

The preliminary results of the methods-test suggest that there is indeed some tendency to pass over such members of the labor force as defined by CPS, although the resulting undercount of the labor force would seem to be only about 0.5 percent. Naturally, this percent might be substantially higher for certain subgroups and occupations such as, say, babysitters and other low-paid and casual workers.[19]

Since the earlier described partial-rotation pattern on which the CPS sample design is founded conveys to this sample some of the aspects of a panel, some of the nonsampling-error-creating effects of panel-type surveys may also be suspected of interfering with CPS estimates.

Among such consequences of panel operations which are likely to distort the statistical results unless careful countermeasures are taken, is the so-called conditioning of the respondent. By this is meant the effect that repetitive interviewing within relatively short intervals may exert on the respondent and on his answer. This effect may extend from complete refusal to minor forms of noncooperation and lack of interest, or to a coloring of the answer during subsequent interviews without the respondent's awareness. It is, of course, also possible that the conditioning effect may in part reflect the interviewer's own reaction to repetitive interviews of the same person.[20] At any rate, a bias may result from the effect, and the analytical user of the data should be aware of its existence and magnitude.

Among the several consequences, the so-called "first-month bias" has been known to exist since the early history of CPS. (No bias in the sense of statistical theory of estimation is implied by the term. "First-month bias" may, of course, result also from factors other than the conditioning effect.) By this "bias" is meant the observed fact that answers of some respondents in households which were interviewed the first time tended to differ from the answers given to the identical questions upon subsequent interview. The result, which is illustrated for employment status measures over the period from March 1959 to December 1961,[21] shows the following percent declines computed from data referring to changes in the status of an identical person (the particular categories were selected because of their relatively high rate of attrition from the first to the last — that is, the eighth interview): 11.5 percent for the unemployed; over 7 percent for private household workers (domestics, babysitters, or jobworkers), and for nonagricultural unpaid family workers; and about 3.5 percent for the nonagricultural self-employed. The corresponding rates from the first to the second interview were: 6.5 percent for the unemployed; 5.5 percent for private household workers; and about 3.5 percent for persons who work only part time, for other than economic reasons.

---

[15] Based on unpublished materials in the files of the Census Bureau and on discussions with Census staff.

[16] Item 21 in the regular CPS form.

[17] Item 23 in the regular CPS form.

[18] Item 26d in the regular CPS form.

[19] It should be noted that problems of this nature are difficult ones which, as a rule, cannot be simply removed by more intensive probing since such probing is quite likely to introduce an opposite bias. In this particular instance persons would be counted as members of the labor force who should not, and without interview probing would not, be so counted. The situation is vaguely comparable to the indeterminacy problem faced by the physicist. Again, the analyst must use a good deal of imagination and resourcefulness to deal with such situations; simply abstention because here and there a sampling error is too large brings no solution to the problem.

[20] Robert B. Pearl and Joseph Waksberg, *The Effect of Repeated Household Interviews in the Current Population Survey,* paper presented before the 47th National Conference of the American Marketing Association, Dallas, June 1964; and Robert B. Pearl, "Gross Changes in the Labor Force: A Problem in Statistical Measurement," *in Employment and Earnings,* April 1963.

[21] Based on data kindly made available by the Census Bureau.

From the analyst's point of view, these first-month biases may or may not be of particular importance, depending on the objective of the analysis.

There is, however, a whole class of interesting and potentially very rewarding problems, the answers to which would be made considerably easier if approached by a longitudinal analysis of the cohort type. Also in this category belongs the so-called gross change analysis to be mentioned later when discussing data uses. In such instances it is, of course, important that the analyst be able to form an opinion as to whether individuals remained from one period to another in the same, say, labor-force-status category or whether they moved into another; and if so, whether the information yields a "stayer" or a "mover" because of a reporting error. Here relatively small percentage differences may be of critical importance, especially if the analysis is one of cohorts which are small compared with the size of the total population; similarly, the problem becomes acute where small intermonthly changes may have to carry the brunt of an analytical argument concerning the pattern of behavior of groups subject to large "first-month bias."

The Current Population Survey Reinterview Program has been in operation for a longer period than most other test programs and experiments, and has had more time to produce interesting results. Its basic approach is to project CPS results, that is, the original interview against reinterview as a first step in getting at nonsampling errors. On sufficiently high levels of aggregation, the results of the program would seem to suggest that nonsampling errors did not introduce a disturbing level of bias in most instances. For example, for the period from January 1955 to June 1961 a study of so-called "reinterview failures"[22] suggests that content failures during the period under consideration were found in about 2.5 percent of the total number of all reinterview assignments. Nevertheless, problems do arise as one moves to lower levels of aggregation and to smaller taxonomic classes.

Although the results of reinterview and other methods-tests undertaken by the Census Bureau are by no means complete and conclusive, they should prove extremely useful to the analyst. Efforts are now underway to add to this store of knowledge and experience on a continuing basis.

Some of the latest developments are now briefly touched upon, not because they have resulted in a great deal of information on nonsampling errors to date but because of their approach and the promise they bear.

Among such efforts of particular interest to the analyst — even though they are concerned primarily with checking population census data — are the attempts by the staff of the Bureau of the Census to develop a measure of "bias"— that is, of difference between a survey-based estimate and the corresponding true value. This measure, which is one of the "accuracy"[23] of the specific survey results, is based at this time on comparisons of census with CPS survey results; it is based on what is referred to as "intensive interview" for a sample of just below 10,000, and on records checks for such items as age, sex, occupation, and industry.

There are also intriguing attempts underway to quantify what might be referred to as the "precision" of the survey results as opposed to the just-discussed accuracy. The measure proposed at this time is the ratio of "simple response variability"— that is, the variability of the individual response deviations "over all possible trials" to the maximum value of the response variability, i.e., the sampling variance of a sample of size one.

This variance ratio, which varies inversely as what might be called the reliability of a particular taxonomic classification, has been given the name of "index of inconsistency."[24]

---

[22] That is, an excessive number of errors in the original interview as revealed by reinterview — excessive in terms of accepted tolerance limits — for which see U.S. Bureau of the Census, ***The Current Population Survey Reinterview Program: Some Notes and Discussion,*** Technical Paper No. 6, 1963, p. 34.

[23] The term "precision" in statistical practice usually refers to sampling error, i.e., to the expected difference between the result of a sample survey and what one would have obtained if the survey had been a complete census using otherwise the same measurements, survey methods, and approach. In other words, precision reflects the difference between sample estimate and a corresponding estimate without any sampling error. "Accuracy," on the other hand, relates to the difference between sample result and true value. It therefore depends on a variety of sometimes not easily or not at all measurable elements. In other words, precision is what can be measured provided the rules of probability selection have been followed; accuracy, however, is what is usually of prime importance to the analyst, and it is the property about which he must form an opinion when interpreting the data.

[24] The model and building stones for these measures are based on the concept of a mean square error of the proportion of individuals classified into a particular taxonomic class. It has two components: a total variance, and a bias component. The total variance, in turn, is divided into response variance, sampling variance, and interaction. The response variance can be thought of as the average of the response deviations for a given sample. One of the obvious difficulties is to evaluate these components in terms of

(Continued on page 59)

Being in the nature of a variance ratio, this index is a pure number. In practice, the analyst may interpret it as suggesting that there is less and less information the higher the numerical value of the index of inconsistency; in other words, the larger this ratio, the higher the suspected nonsampling error. Like the coefficient of determination, the square of the correlation coefficient which it resembles, this ratio is not linear. Experimental work to date, and attempts to evaluate this index of inconsistency in specific instances would suggest that numerical values for this index in excess of .5 describe a highly unsatisfactory situation with respect to the presence of nonsampling error.

Thus, the practicing analyst has here a quantifiable "handle" by which to get hold of the nonsampling error situation quickly in specific instances; in spite of its limitations, and although the statistical models underlying this measure are conceptually by no means simple and trivial, its practical usefulness when applied on a "go-no-go" basis (such as the above-suggested dividing line at .5 or .6) is obvious. In addition, it enables the analyst, when confronted with a choice of variables, to rank them in the order of their suspected nonsampling error as expressed by this index. Hence, the decision by the analyst concerning nonsampling error is made considerably easier without the necessity of his having recourse to detailed and intensive case study.

So far, the test program here discussed is aimed at ascertaining the presence and pattern of nonsampling errors in the decennial census. Although the concepts and methods developed under this program are also pertinent to other kinds of survey operations, the transfer from one type of survey to another cannot be undertaken automatically.

With respect to CPS, the systematic and empirical study of enumeration errors has a history of more than 15 years. In general, the attempts to control the quality of survey results have been based on reinterview of a sample of original interviewees. Since early in 1955, qualified personnel have used these reinterviews as checks on a sample of each interviewer's work at the approximate rate of once every four months, i.e., three times per year. This procedure has resulted in a reinterview of about 7.5 percent of the CPS households each month.[25] The reinterview now covers all questions pertaining to work status in addition to a miscellany of questions which vary over time. Thus, this reinterview system is in essence a statistical quality-control program aimed at establishing whether interview results meet certain quality-control standards for each of three major interview aspects: coverage of sampling units, coverage of persons, and content of interview. The findings from these tests have been carefully recorded,[26] and they are being carried forward in internal Census memoranda, the so-called evaluation reports series.

Of direct interest to the labor force analyst who is concerned about nonsampling errors are those parts of the CPS Reinterview Program which correspond to the above-mentioned index of inconsistency. Here the relevant measure is the ratio of what is referred to as the gross difference rate (g) to 2pq,[27] the approximate maximum value of the gross difference rate between two surveys. This ratio $\frac{g}{(2pq)}$ can be interpreted by analogy to the earlier mentioned index of inconsistency, keeping in mind that the ratio g to 2pq is 100 times the numerical value of the index.

At this time, the numerical evaluations of this ratio are published for relatively few major labor-force-status groups,[28] but the test program is continuing, and it is expected that the results will again be presented and before long be brought up-to-date in a revision of *The Current Population Survey Reinterview Program.* Also, there are available at this time gross difference rate estimates for a few more variables than are now published.

In conclusion, it may be said that in the occasional practice of data-consumers and of some well-meaning data-producers, perhaps too much is made of the sampling errors and too little of the nonsampling errors. Be that as it may, the ability to evaluate and qualify empirical data is a skill of great consequence for successful analysis.

As to sampling errors, acquiring a feeling for them and forming an opinion about their general and complex nature and order of magnitude are frequently much more useful than automatic acceptance or rejection of available data on the basis of mechanical rules. Rarely, if ever, is the analyst in this area of research confronted with a problem situation akin to that facing, say, the statistical quality-control engineer; and rarely is the manpower analyst justified in the use of hard-and-fast rules which, most often, apply to quite different circumstances in which risk, cost of an error, and the like can be evaluated with near actuarial precision.

---

unbiased estimates from specific survey results. See Morris H. Hansen, William N. Hurwitz, and Leon Pritzker, *The Estimation and Interpretation of Close Differences and the Simple Response Variance,* manuscript prepared for publication in the forthcoming 70th Birthday Volume in honor of Professor P. C. Mahalanobis; Leon Pritzker and Robert Hanson, "Measurement Errors in the 1960 Census of Population," *1962 Proceedings of the Social Statistics Section of the American Statistical Associaion;* and Morris H. Hansen, William N. Hurwitz, and Max A. Bershad, *Measurement Errors in Censuses and Surveys,* a paper presented at the Annual Meeting of the International Statistical Institute, 32nd Session, Tokyo, 1960.

[25] Census, *The Current Population Survey Reinterview Program.*

[26] *Ibid.,* Appendix B.

[27] *Ibid.,* pp. 9 ff.; p is the average percent in the two surveys, the original and the reinterview one, having the given characteristic; q equals 1 minus p.

[28] *Ibid.,* table E, p. 13.

The two nomograms presented in the text, approximate and rough as they may be, should provide the analyst with a simple tool for the qualification of the data where sampling errors are the issue. A feeling for such sampling errors may be developed by actual work experience with specific data, as well as by familiarity with the theory and art of sampling. For multistage designs such as the one used by CPS, an example of such familiarity might be the acquisition of a consciousness of, and a feeling for, intraclass correlation. Even a speculative answer to the question of whether, on a given level of aggregation, intraclass correlation is likely to be higher than average for a particular variable or characteristic can lend a good deal of support to the analyst's decision of whether to assume that the actual sampling error is likely to be considerably larger or smaller than the one suggested by straightforward sampling variances such as those on which the nomograms in figures 1 and 2 are based.

More disturbing is the problem of nonsampling errors. As far as CPS is concerned, considerable and promising research has been, and is continuing to be, undertaken. However, many of the results of practical importance to the analyst are yet to come. Therefore, the forward-looking analyst had better keep in close touch with current work by the Census Bureau in the as-yet-little-explored area of nonsampling errors.

Appendix III.

Current Population Survey Schedules and Forms

From the analyst's point of view, the most useful data inventory would probably be a table of contents generated by an appropriate data-storage and -retrieval system to describe its own content. In the absence of such an inventory at this time, it would seem that the CPS schedules provide the best source of information on what kind of information exists — in particular because virtually all data gathered on the basic CPS schedules are now recorded on magnetic tape.

Several sets of CPS schedules are in use. There is the "regular" schedule which is administered to all sample respondents, once every month; there is the "control card" which is used for every household in the sample but only once, i.e., at the time the household appears for the first time in the sample; and there are the many "supplemental" schedules which range over a great variety of subjects, from recreational patterns, immunization, and veterans to the famous quarterly survey if intentions, i.e., of consumers' buying plans for houses, automobiles, and other durable goods. Supplemental questions may be addressed to the full CPS sample or only to a subsample of the households.

The material in this appendix is limited to schedules which would seem to be of most direct interest in manpower analysis:

1. The basic CPS schedule — the so-called FOSDIC schedule.
2. The control card.
3. The supplemental questions on work experience, essentially as used since 1956, and administered to the full CPS sample. This supplement is also known as the February supplement.
4. The annual survey on income, family and marital characteristics, migration, and educational attainment. This survey is administered to the entire CPS sample with the exception of the income question which is given to a subsample of 3 out of 4. This survey is also known as the March supplement.
5. The survey of dual-jobholders and of premium pay which is directed to the full CPS sample, and is also known as the May supplement.
6. The survey on reasons for looking for work undertaken simultaneously with the survey on fertility and child spacing. This survey, which covers the entire CPS sample, is also known as the June supplement.
7. The school enrollment supplement which can be traced as far back as 1947 and which is also known as the October supplement. It secures for the entire CPS sample school enrollment and related information on civilian household members 5 to 34 years old, and also information on dropouts.
8. The basic CPS schedules in use from February 1952 to October 1961 (8a is the form administered to household heads, and 8b the form for other household members, 14 years of age or older.)

The basic CPS schedule referred to under (1) above has been in use, essentially in the form shown, since October 1961. Prior to that date and starting with February 1952 a somewhat shorter schedule was in use (8 above) which is reproduced to convey a general idea of the kind of information collected during that earlier period. In interpreting this information, it must be kept in mind that — as mentioned earlier — machine readable data are not likely to be found readily, if at all, for the years prior to 1959.

1. BASIC OR *FOSDIC* SCHEDULE

Jan Mar May Jul Sep Nov Feb Apr Jun Aug Oct Dec

○

Form Approved: Budget Bureau No. 41-R 1202.13

FORM CPS-1
(4-30-65)

U.S. DEPARTMENT OF COMMERCE
BUREAU OF THE CENSUS

CURRENT POPULATION SURVEY
JULY 1965

2. CONTROL NUMBER

3. SAMPLE

A

1. INTERVIEWER CHECK ITEM

Only CPS-1 for household .... ○
First CPS-1 of continuation h'hld . ○
(Fill all applicable items on this page)

Second CPS-1 of continuation h'hld ○
Third, fourth, etc. CPS-1 .... ○
(Transcribe items 2 through 16 from first CPS-1)

4a. *(Transcribe from Control Card Items 10a, 11, or 12)*

A	B	C	D	E
○	○	○	○	○

(If no code, mark acreage in item 4b)

4b. *Transcribe from control card item 10b if item 4a is blank*

10 acres or more ○
Less than 10 acres ○

5. TYPE OF LIVING QUARTERS

HOUSING UNIT

House, apartment, flat .... ○
HU in nontransient hotel, motel, etc. ○
HU, permanent, in transient hotel, motel, etc. .... ○
HU in rooming housing .... ○
Trailer, permanent .... ○
Trailer, mobile .... ○
HU not specified above .... ○
(Describe below)

OTHER UNIT

Quarters not HU in rooming or boarding house .... ○
Unit not permanent in transient hotel, motel, etc. ○
Tent site or trailer site .... ○
Other not HU .... ○
(Describe below).

6. PSU NO.	7. SERIAL NO.	8. IDEN CODE	9. HOUSE HOLD NO.	10. SEGMENT NO.
0 ○ ○	0 ○ ○ ○			0 ○ ○ ○
1 ○ ○ ○	1 ○ ○ ○ ○	1 ○ ○	1 ○	1 ○ ○ ○
2 ○ ○ ○	2 ○ ○ ○ ○	2 ○ ○	2 ○	2 ○ ○ ○
3 ○ ○ ○	3 ○ ○ ○ ○	3 ○ ○	3 ○	3 ○ ○ ○
4 ○ ○ ○	4 ○ ○ ○ ○	4 ○ ○	4 ○	4 ○ ○ ○
5 ○ ○ ○	5 ○ ○ ○ ○	5 ○	5 ○	5 ○ ○ ○
6 ○ ○ ○	6 ○ ○ ○ ○	6 ○	6 ○	6 ○ ○ ○
7 ○ ○ ○	7 ○ ○ ○ ○	7 ○	7 ○	7 ○ ○ ○
8 ○ ○ ○	8 ○ ○ ○ ○	8 ○		8 ○ ○ ○
9 ○ ○ ○	9 ○ ○ ○			9 ○ ○ ○

FILL FOR SPECIAL DWELLING PLACE

11a. NAME OF PLACE

11b. TYPE OF PLACE CODE

12a. SAMPLE UNIT NO.

12b. DESCRIPTION OF SAMPLE UNIT (Room No., Bed No., etc.)

INTERVIEW DATA

13. INTERVIEWER CODE

A	B	C	D	E	F	G	H	J	K
○	○	○	○	○	○	○	○	○	○

1	2	3	4	5	6	7	8	9	10	11
○	○	○	○	○	○	○	○	○	○	○

14. DATE COMP.

15. LINE NO. H'HLD RESP

1 2 3 4 5 6

16a. TYPE INTERVIEW

Personal ..... ○
Tel.-Callback . ○
ICR filled .... ○
Noninterview . . ○
(Go to 16b)

Tel.-Regular . . ○
(Fill 16c and 16d)

16b. NO. PERS. VISITS
(Exc. Reg. Tel.)

1 ○
2 ○
3+ ○

(If Nonint., fill 17. If Int., fill 16c-d)

16c. TIME OF INTERVIEW
(For all cases except noninterviews)

Before noon. ○ 5 p.m.-7 p.m. ○
Noon-5 p.m. ○ After 7 p.m. ○

16d. DAY OF INTERVIEW

M	T	W	T	F	S	S	M	T
○	○	○	○	○	○	○	○	○

17. NONINTERVIEW *(Mark one reason in Item 17a or 17b. If Item 17a marked, fill Race of Head. If Item 17b marked, fill Items 17c-e, as applicable.)*

17a. TYPE A

(Mark noninterview reason and race of head below.)

REASON

No one home ○
Temporarily absent .... ○
Refused .... ○
Other-Occ. . . ○
(Describe below)

RACE OF HEAD

White ..... ○
Negro .... ○
Other ..... ○

17b. TYPE B OR C

TYPE B

Vacant - regular .... ○
Vacant - storage of h'hld furniture .... ○
Temp. occ. by persons with URE .... ○
(Fill 17c)

Unfit or to be demolished .... ○
Under construction, not ready .... ○
Converted to temp. business or storage .... ○
Occ. by Armed Force members or persons under 14 ○
Unoccupied tent site or trailer site .... ○
Permit granted, construction not started .... ○

TYPE C

Demolished .... ○
House or trailer moved .... ○
Outside segment .... ○
Converted to permanent business or storage ..... ○
Merged .... ○
Condemned .... ○
Built after April 1, 1960 .... ○

Other *(Specify below)* .... ○
(Omit 17c-e)

17c. SEASONAL STATUS

Year round .... ○ *(Fill HVS if HU in Item 5)*

Migratory workers . . . ○
Seasonal .... ○
(Fill Items 17d and 17e below if HU in Item 5)

17d. Is this unit usually occupied:

Summers only ..... ○
Winters only ...... ○
Other .... ○
(Describe below)

(Transcribe to back of Control Card first time unit is vacant during a 4-month enumeration period.)

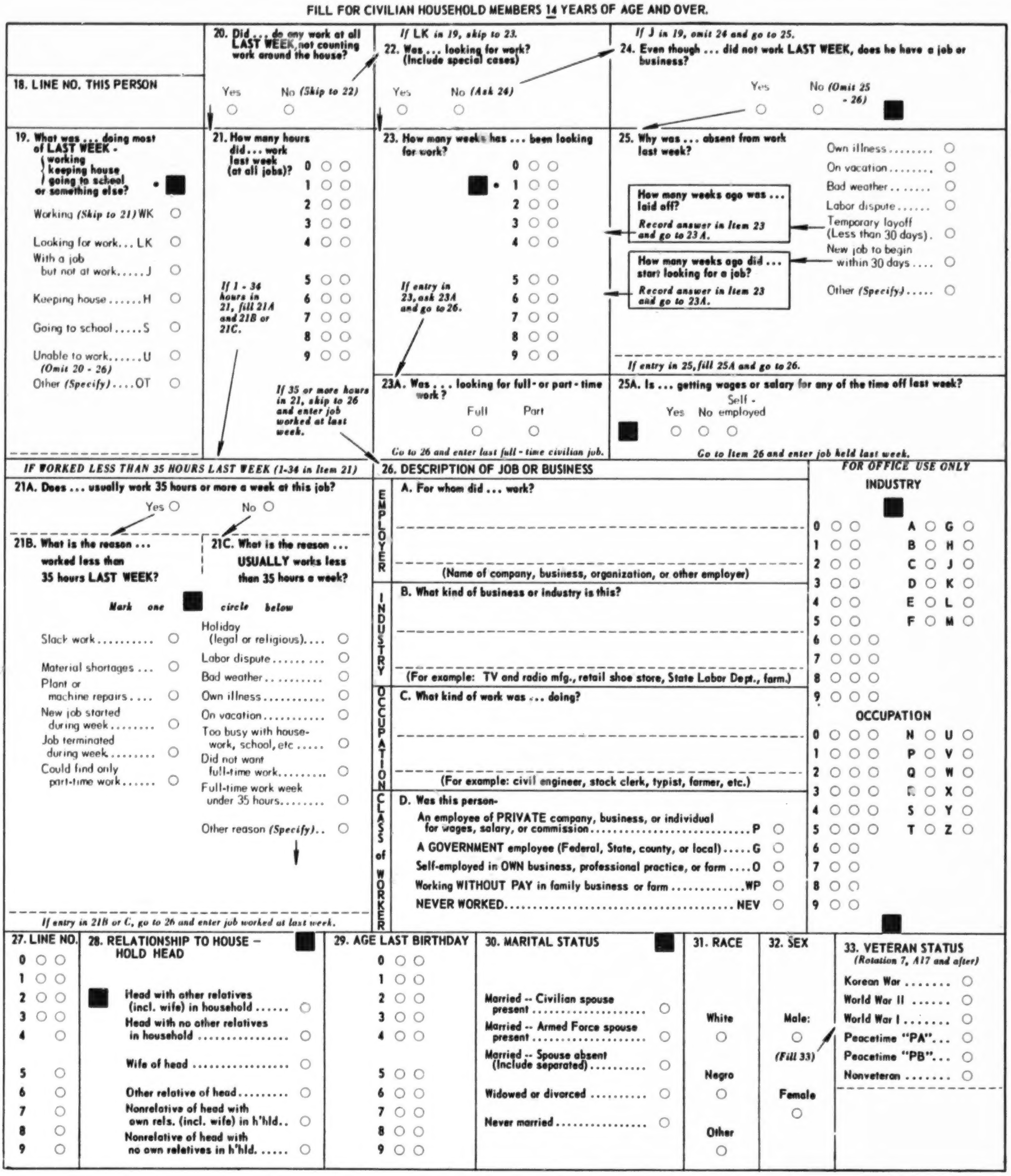
17.e. CONDITION OF UNIT

Sound .... ○
Deteriorating ...... ○
Dilapidated .... ○

(Transcribe from back of Control Card in each succeeding month unit is vacant in the same 4-month enumeration period.)

NOTES:

1. BASIC OR *FOSDIC* SCHEDULE *(continued)*



FILL FOR CIVILIAN HOUSEHOLD MEMBERS 14 YEARS OF AGE AND OVER.

18. LINE NO. THIS PERSON

20. Did ... do any work at all LAST WEEK, not counting work around the house?

Yes ○ No *(Skip to 22)* ○

If LK in 19, skip to 23.

22. Was ... looking for work? (Include special cases)

Yes ○ No *(Ask 24)* ○

If J in 19, omit 24 and go to 25.

24. Even though ... did not work LAST WEEK, does he have a job or business?

Yes ○ No *(Omit 25 - 26)* ○

19. What was ... doing most of LAST WEEK - working, keeping house, going to school or something else?

- Working *(Skip to 21)* WK ○
- Looking for work... LK ○
- With a job but not at work..... J ○
- Keeping house...... H ○
- Going to school..... S ○
- Unable to work...... U *(Omit 20 - 26)* ○
- Other *(Specify)*.... OT ○

21. How many hours did ... work last week (at all jobs)?

0 ○ ○
1 ○ ○
2 ○ ○
3 ○ ○
4 ○ ○
5 ○ ○
6 ○ ○
7 ○ ○
8 ○ ○
9 ○ ○

If 1 - 34 hours in 21, fill 21A and 21B or 21C.

If 35 or more hours in 21, skip to 26 and enter job worked at last week.

23. How many weeks has ... been looking for work?

0 ○ ○
1 ○ ○
2 ○ ○
3 ○ ○
4 ○ ○
5 ○ ○
6 ○ ○
7 ○ ○
8 ○ ○
9 ○ ○

If entry in 23, ask 23A and go to 26.

23A. Was . . . looking for full- or part-time work?

Full ○ Part ○

Go to 26 and enter last full-time civilian job.

25. Why was ... absent from work last week?

- Own illness........ ○
- On vacation........ ○
- Bad weather....... ○
- Labor dispute...... ○
- Temporary layoff (Less than 30 days). ○
- New job to begin within 30 days.... ○
- Other *(Specify)*..... ○

How many weeks ago was ... laid off?
Record answer in Item 23 and go to 23 A.

How many weeks ago did ... start looking for a job?
Record answer in Item 23 and go to 23A.

If entry in 25, fill 25A and go to 26.

25A. Is ... getting wages or salary for any of the time off last week?

Yes ○ No ○ Self-employed ○

Go to Item 26 and enter job held last week.

IF WORKED LESS THAN 35 HOURS LAST WEEK (1-34 in Item 21)

21A. Does ... usually work 35 hours or more a week at this job?

Yes ○ No ○

21B. What is the reason ... worked less than 35 hours LAST WEEK?

21C. What is the reason ... USUALLY works less than 35 hours a week?

Mark one circle below

- Slack work.......... ○
- Material shortages ... ○
- Plant or machine repairs.... ○
- New job started during week........ ○
- Job terminated during week........ ○
- Could find only part-time work...... ○
- Holiday (legal or religious).... ○
- Labor dispute......... ○
- Bad weather.......... ○
- Own illness........... ○
- On vacation........... ○
- Too busy with housework, school, etc..... ○
- Did not want full-time work......... ○
- Full-time work week under 35 hours........ ○
- Other reason *(Specify)*.. ○

If entry in 21B or C, go to 26 and enter job worked at last week.

26. DESCRIPTION OF JOB OR BUSINESS

EMPLOYER

A. For whom did ... work?

(Name of company, business, organization, or other employer)

INDUSTRY

B. What kind of business or industry is this?

(For example: TV and radio mfg., retail shoe store, State Labor Dept., farm.)

OCCUPATION

C. What kind of work was ... doing?

(For example: civil engineer, stock clerk, typist, farmer, etc.)

CLASS of WORKER

D. Was this person-

- An employee of PRIVATE company, business, or individual for wages, salary, or commission.......... P ○
- A GOVERNMENT employee (Federal, State, county, or local)..... G ○
- Self-employed in OWN business, professional practice, or farm .... O ○
- Working WITHOUT PAY in family business or farm .... WP ○
- NEVER WORKED.......... NEV ○

FOR OFFICE USE ONLY

INDUSTRY

0 ○ ○ A ○ G ○
1 ○ ○ B ○ H ○
2 ○ ○ C ○ J ○
3 ○ ○ D ○ K ○
4 ○ ○ E ○ L ○
5 ○ ○ F ○ M ○
6 ○ ○ ○
7 ○ ○ ○
8 ○ ○ ○
9 ○ ○ ○

OCCUPATION

0 ○ ○ ○ N ○ U ○
1 ○ ○ ○ P ○ V ○
2 ○ ○ ○ Q ○ W ○
3 ○ ○ ○ R ○ X ○
4 ○ ○ ○ S ○ Y ○
5 ○ ○ ○ T ○ Z ○
6 ○ ○
7 ○ ○
8 ○ ○
9 ○ ○

27. LINE NO.

0 ○ ○
1 ○ ○
2 ○ ○
3 ○ ○
4 ○
5 ○
6 ○
7 ○
8 ○
9 ○

28. RELATIONSHIP TO HOUSEHOLD HEAD

- Head with other relatives (incl. wife) in household...... ○
- Head with no other relatives in household .... ○
- Wife of head .... ○
- Other relative of head......... ○
- Nonrelative of head with own rels. (incl. wife) in h'hld.. ○
- Nonrelative of head with no own relatives in h'hld. ..... ○

29. AGE LAST BIRTHDAY

0 ○ ○
1 ○ ○
2 ○ ○
3 ○ ○
4 ○ ○
5 ○ ○
6 ○ ○
7 ○ ○
8 ○ ○
9 ○ ○

30. MARITAL STATUS

- Married -- Civilian spouse present .... ○
- Married -- Armed Force spouse present .... ○
- Married -- Spouse absent (Include separated).......... ○
- Widowed or divorced .... ○
- Never married................ ○

31. RACE

- White ○
- Negro ○
- Other ○

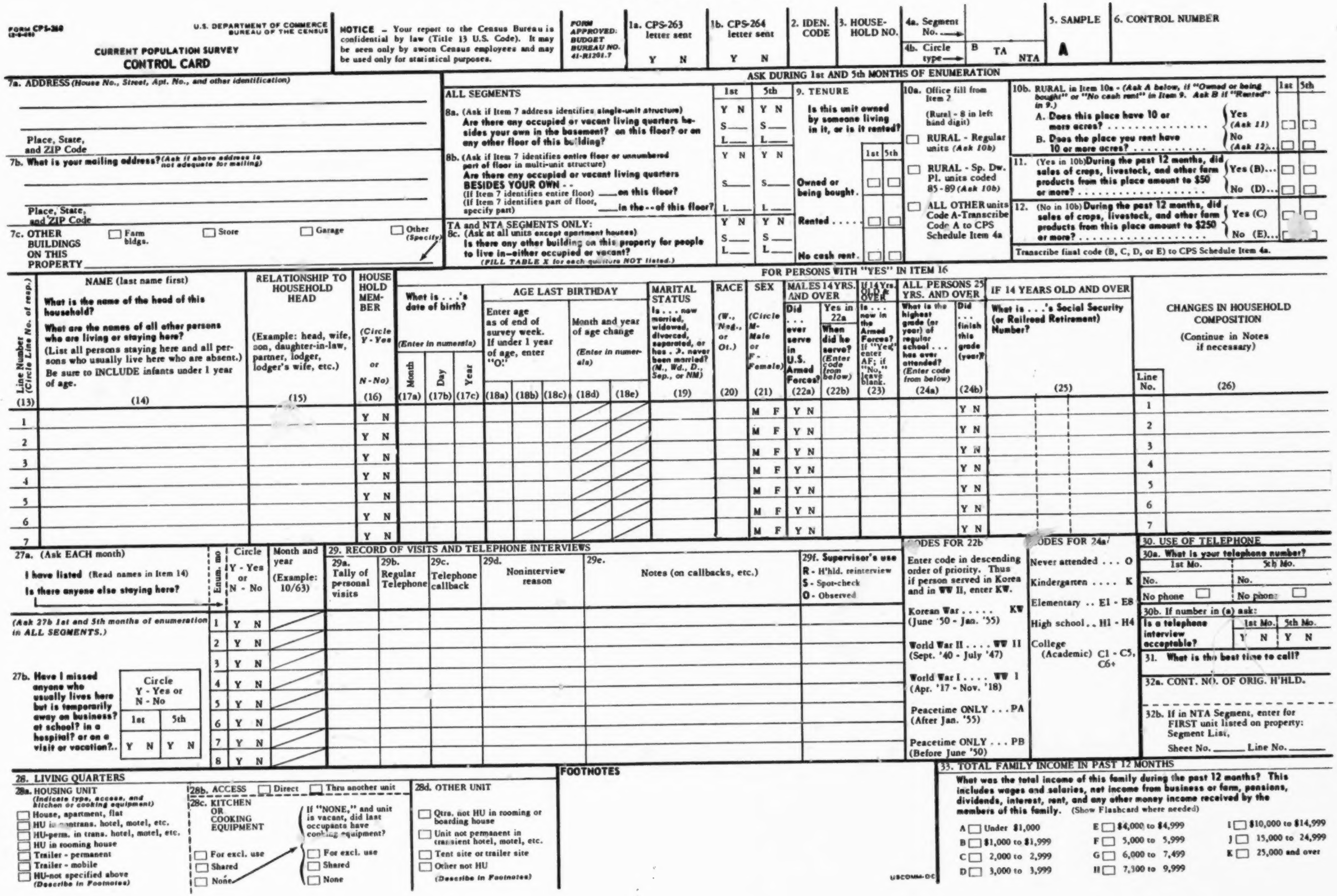
32. SEX

- Male: ○ *(Fill 33)*
- Female ○

33. VETERAN STATUS *(Rotation 7, A17 and after)*

- Korean War .... ○
- World War II ...... ○
- World War I .... ○
- Peacetime "PA"... ○
- Peacetime "PB"... ○
- Nonveteran .... ○

2. CONTROL CARD



FORM CPS-260

U.S. DEPARTMENT OF COMMERCE
BUREAU OF THE CENSUS

CURRENT POPULATION SURVEY
CONTROL CARD

NOTICE – Your report to the Census Bureau is confidential by law (Title 13 U.S. Code). It may be seen only by sworn Census employees and may be used only for statistical purposes.

FORM APPROVED: BUDGET BUREAU NO. 41-R1201.7

1a. CPS-263 letter sent	1b. CPS-264 letter sent	2. IDEN. CODE	3. HOUSE-HOLD NO.	4a. Segment No. → / 4b. Circle type →		5. SAMPLE	6. CONTROL NUMBER
Y N	Y N				B TA NTA	A	

7a. ADDRESS *(House No., Street, Apt. No., and other identification)*

Place, State, and ZIP Code

7b. **What is your mailing address?** *(Ask if above address is not adequate for mailing)*

Place, State, and ZIP Code

7c. OTHER BUILDINGS ON THIS PROPERTY ☐ Farm bldgs. ☐ Store ☐ Garage ☐ Other *(Specify)*

ASK DURING 1st AND 5th MONTHS OF ENUMERATION

ALL SEGMENTS	1st	5th
8a. (Ask if Item 7 address identifies single-unit structure) **Are there any occupied or vacant living quarters besides your own in the basement? on this floor? or on any other floor of this building?**	Y N S___ L___	Y N S___ L___
8b. (Ask if Item 7 identifies entire floor or unnumbered part of floor in multi-unit structure) **Are there any occupied or vacant living quarters BESIDES YOUR OWN - -** (If Item 7 identifies entire floor) ___**on this floor?** (If Item 7 identifies part of floor, specify part) ___**in the - - of this floor?**	Y N S___ L___	Y N S___ L___
TA and NTA SEGMENTS ONLY: 8c. (Ask at all units except apartment houses) **Is there any other building on this property for people to live in—either occupied or vacant?** *(FILL TABLE X for each quarters NOT listed.)*	Y N S___ L___	Y N S___ L___

9. TENURE

Is this unit owned by someone living in it, or is it rented?

	1st	5th
Owned or being bought .	☐	☐
Rented	☐	☐
No cash rent .	☐	☐

10a. Office fill from Item 2

(Rural - 8 in left hand digit)

☐ RURAL - Regular units *(Ask 10b)*

☐ RURAL - Sp. Dw. Pl. units coded 85 - 89 *(Ask 10b)*

☐ ALL OTHER units Code A-Transcribe Code A to CPS Schedule Item 4a

		1st	5th
10b. RURAL in Item 10a - *(Ask A below, if "Owned or being bought" or "No cash rent" in Item 9. Ask B if "Rented" in 9.)* A. **Does this place have 10 or more acres?** . . . . B. **Does the place you rent have 10 or more acres?**	Yes *(Ask 11)*	☐	☐
	No *(Ask 12)* . .	☐	☐
11. (Yes in 10b) **During the past 12 months, did sales of crops, livestock, and other farm products from this place amount to $50 or more?**	Yes (B) . . .	☐	☐
	No (D) . . .	☐	☐
12. (No in 10b) **During the past 12 months, did sales of crops, livestock, and other farm products from this place amount to $250 or more?**	Yes (C)	☐	☐
	No (E) . . .	☐	☐

Transcribe final code (B, C, D, or E) to CPS Schedule Item 4a.

FOR PERSONS WITH "YES" IN ITEM 16

Line Number *(Circle Line No. of resp.)*	NAME (last name first) **What is the name of the head of this household?** **What are the names of all other persons who are living or staying here?** (List all persons staying here and all persons who usually live here who are absent.) Be sure to INCLUDE infants under 1 year of age.	RELATIONSHIP TO HOUSEHOLD HEAD (Example: head, wife, son, daughter-in-law, partner, lodger, lodger's wife, etc.)	HOUSE HOLD MEMBER *(Circle Y - Yes or N - No)*	**What is . . .'s date of birth?** *(Enter in numerals)* Month	Day	Year	AGE LAST BIRTHDAY: Enter age as of end of survey week. If under 1 year of age, enter "O."			Month and year of age change *(Enter in numerals)*		MARITAL STATUS **Is . . . now married, widowed, divorced, separated, or has . . . never been married?** *(M., Wd., D., Sep., or NM)*	RACE *(W., Neg., or Ot.)*	SEX *(Circle M-Male or F-Female)*	MALES 14 YRS. AND OVER: **Did . . . ever serve in U.S. Armed Forces?**	Yes in 22a: **When did he serve?** *(Enter code from below)*	If 14 Yrs. OLD & OVER: **Is . . . now in the Armed Forces?** If "Yes" enter AF; if "No," leave blank.	ALL PERSONS 25 YRS. AND OVER: **What is the highest grade (or year) of regular school . . . has ever attended?** *(Enter code from below)*	**Did . . . finish this grade (year)?**	IF 14 YEARS OLD AND OVER: **What is . . .'s Social Security (or Railroad Retirement) Number?**	Line No.	CHANGES IN HOUSEHOLD COMPOSITION (Continue in Notes if necessary)
(13)	(14)	(15)	(16)	(17a)	(17b)	(17c)	(18a)	(18b)	(18c)	(18d)	(18e)	(19)	(20)	(21)	(22a)	(22b)	(23)	(24a)	(24b)	(25)		(26)
1			Y N											M F	Y N				Y N		1	
2			Y N											M F	Y N				Y N		2	
3			Y N											M F	Y N				Y N		3	
4			Y N											M F	Y N				Y N		4	
5			Y N											M F	Y N				Y N		5	
6			Y N											M F	Y N				Y N		6	
7			Y N											M F	Y N				Y N		7	

27a. (Ask EACH month)

I have listed (Read names in Item 14)

Is there anyone else staying here?

(Ask 27b 1st and 5th months of enumeration in ALL SEGMENTS.)

27b. **Have I missed anyone who usually lives here but is temporarily away on business? at school? in a hospital? or on a visit or vacation?** . .

Circle Y - Yes or N - No	
1st	5th
Y N	Y N

29. RECORD OF VISITS AND TELEPHONE INTERVIEWS

Enum. mo	Circle Y - Yes or N - No	Month and year (Example: 10/63)	29a. Tally of personal visits	29b. Regular Telephone	29c. Telephone callback	29d. Noninterview reason	29e. Notes (on callbacks, etc.)	29f. **Supervisor's use** R - H'hld. reinterview S - Spot-check O - Observed
1	Y N							
2	Y N							
3	Y N							
4	Y N							
5	Y N							
6	Y N							
7	Y N							
8	Y N							

CODES FOR 22b

Enter code in descending order of priority. Thus if person served in Korea and in WW II, enter KW.

Korean War . . . . . KW
(June '50 - Jan. '55)

World War II . . . . WW II
(Sept. '40 - July '47)

World War I . . . . WW I
(Apr. '17 - Nov. '18)

Peacetime ONLY . . . PA
(After Jan. '55)

Peacetime ONLY . . . PB
(Before June '50)

CODES FOR 24a

Never attended . . . O

Kindergarten . . . . K

Elementary . . E1 - E8

High school . . H1 - H4

College (Academic) C1 - C5, C6+

30. USE OF TELEPHONE

30a. **What is your telephone number?**

1st Mo.	5th Mo.
No.	No.
No phone ☐	No phone ☐

30b. If number in (a) ask:

	1st Mo.	5th Mo.
Is a telephone interview acceptable?	Y N	Y N

31. **What is the best time to call?**

32a. CONT. NO. OF ORIG. H'HLD.

32b. If in NTA Segment, enter for FIRST unit listed on property: Segment List,
Sheet No. ______ Line No. ______

28. LIVING QUARTERS

28a. HOUSING UNIT *(Indicate type, access, and kitchen or cooking equipment)*

- ☐ House, apartment, flat
- ☐ HU in nontrans. hotel, motel, etc.
- ☐ HU-perm. in trans. hotel, motel, etc.
- ☐ HU in rooming house
- ☐ Trailer - permanent
- ☐ Trailer - mobile
- ☐ HU-not specified above *(Describe in Footnotes)*

28b. ACCESS ☐ Direct ☐ Thru another unit

28c. KITCHEN OR COOKING EQUIPMENT

- ☐ For excl. use
- ☐ Shared
- ☐ None

If "NONE," and unit is vacant, did last occupants have cooking equipment?

- ☐ For excl. use
- ☐ Shared
- ☐ None

28d. OTHER UNIT

- ☐ Qtrs. not HU in rooming or boarding house
- ☐ Unit not permanent in transient hotel, motel, etc.
- ☐ Tent site or trailer site
- ☐ Other not HU *(Describe in Footnotes)*

FOOTNOTES

USCOMM-DC

33. TOTAL FAMILY INCOME IN PAST 12 MONTHS

What was the total income of this family during the past 12 months? This includes wages and salaries, net income from business or farm, pensions, dividends, interest, rent, and any other money income received by the members of this family. (Show Flashcard where needed)

A ☐ Under $1,000	E ☐ $4,000 to $4,999	I ☐ $10,000 to $14,999
B ☐ $1,000 to $1,999	F ☐ 5,000 to 5,999	J ☐ 15,000 to 24,999
C ☐ 2,000 to 2,999	G ☐ 6,000 to 7,499	K ☐ 25,000 and over
D ☐ 3,000 to 3,999	H ☐ 7,500 to 9,999	

2. CONTROL CARD *(continued)*

FOR UNITS IN HOUSING VACANCY SURVEY (HVS)

Transcribe from HVS schedule the first month unit is vacant. Transcribe to HVS in subsequent months unit is vacant during same 4-month period of enumeration. If unit is vacant in both the first and second 4-month periods of enumeration, repeat above procedure for second 4-month period of enumeration.

YEAR BUILT *(HVS Item 12)*		NUMBER HOUSING UNITS IN STRUCTURE *(HVS Item 13)*			CONDITION OF UNIT *(HVS Item 14)*			NUMBER OF ROOMS IN UNIT: TOTAL *(HVS Item 15)*			NUMBER OF ROOMS IN UNIT: BEDROOMS *(HVS Item 16)*			PLUMBING FACILITIES FOR THIS UNIT: RUNNING WATER *(HVS Item 17)*			PLUMBING FACILITIES FOR THIS UNIT: FLUSH TOILET *(HVS Item 18)*		
1960 or later			1st period of enum.	2nd period of enum.		1st period of enum.	2nd period of enum.		1st period of enum.	2nd period of enum.		1st period of enum.	2nd period of enum.		1st period of enum.	2nd period of enum.		1st period of enum.	2nd period of enum.
1955-1959		1			Sound			1			None			Hot and cold in str.			For excl. use		
		2			Deteriorating			2			1						Shared		
1950-1954		3-4			Dilapidated			3			2			Cold only in str.			None		
		5-9						4			3						BATHTUB OR SHOWER *(HVS Item 19)*		
1940-1949		10 or more						5			4 or more			Run. water out. str.				1st period of enum.	2nd period of enum.
1930-1939								6									For excl. use		
1929 or earlier								7						No run. water			Shared		
								8 or more									None		

Name and telephone number of noninterview respondent (including HVS).

FOR VACANT SEASONAL HOUSING UNITS AND FOR VACANT SEASONAL HOUSING UNITS FOR MIGRATORY WORKERS

Transcribe from CPS Items 17d and 17e first month unit is vacant. Transcribe from Control Card to CPS schedule Items 17d and 17e in subsequent months unit is vacant during same 4-month period of enumeration. If unit is seasonally vacant in both the first and second 4-month periods of enumeration, repeat above procedure for second 4-month period of enumeration.

IS THIS UNIT USUALLY OCCUPIED-- *(CPS Item 17d)*			CONDITION OF UNIT *(CPS Item 17e)*		
	1st period of enum.	2nd period of enum.		1st period of enum.	2nd period of enum.
Summers only			Sound		
Winters only			Deteriorating		
Other (Descr. in footn.)			Dilapidated		

TABLE X - LIVING QUARTERS DETERMINATIONS AT LISTED ADDRESS

LINE NO.	CONTROL CARD ITEM NO.	Are these *(specify location)* quarters for more than one group of people?		LOCATION OF UNIT	USE OR CHARACTERISTICS: OCCUPIED — Do the occupants of these *(specify location)* quarters live and eat with any other group of people?		ALL QUARTERS — Do these *(specify location)* quarters have: Direct access from the outside or through a common hall?		A kitchen or cooking equipment for exclusive use?		CLASSIFICATION			IF HU IN B SEGMENT ASK:		REMARKS
		Yes *(Fill one line for each group.)*	No	*(Examples: Basement, 2nd floor)*	Yes	No	Yes	No	Yes	No	Not a separate unit *(Add occupants to this Control Card)*	Fill separate Control Card and enumerate: HU	Other unit	In what year were these *(specify location)* quarters created? *(If 1959 or 1960, also specify "F" if first half or "L" if last half.)*	*(If before July 1960)* What was the name of the household head of these quarters on April 1, 1960?	
(1)	(2)	(3a)	(3b)	(4)	(5a)	(5b)	(6a)	(6b)	(7a)	(7b)	(8)	(9a)	(9b)	(10)	(11)	(12)
1																
2																
3																
4																
5																

FOOTNOTES:

3. FEBRUARY SUPPLEMENT

34. In 1964 how many weeks did . . . work either full-time or part time *(not counting work around the house)*? *(Include paid vacations and paid sick leave.)*

(Enter number of weeks and mark one circle) ☐

None ○ *(Ask Item [illegible])*

1-13 wks ○
14-26 wks ○
27-39 wks ○
40-47 wks ○
48-49 wks ○
(Skip to Item 38)

50-52 wks ○ *(Skip to Item 40)*

35. Even though . . . did not work in 1964, did he spend any time trying to find a job?

Yes ○ *(Ask Items 36-37)*
No ○ *(Skip to Item 37)*

36. How many different weeks was . . . looking for work or on layoff from a job?

1-4 wks ○
5-14 wks ○
15-26 wks ○
27-39 wks ○
40 wks or more ○
(Ask 37)

37. What was the main reason . . . did not work in 1964?

Ill or disabled . . . . ○
Keeping house . . . . ○
School . . . . ○
Could not find work . . . ○
Other *(Specify)* . . . . ○

(Skip to Item 45)

38. You said . . . worked about *(entry in Item 34)* weeks in 1964. How many of the remaining *(52 weeks minus entry in Item 34)* weeks was . . . looking for work or on layoff from a job?

(Enter number of weeks and mark one circle) ☐

None . . . . ○ *(Skip to Item 41)*

1-4 wks . . . . . ○
5-10 wks. . . . . ○
11-14 wks. . . . ○
15-26 wks. . . . ○
27-39 wks . . . ○
40 wks or more ○
(Ask Item 39)

39. Were the ______ weeks . . . was looking for work (or on layoff) all in one stretch?

Yes - 1 stretch . . . ○
No - 2 stretches . . ○
No - 3 + stretches ○
(Skip to Item 41)

40. Did . . . lose any full weeks of work in 1964 because he was on layoff from a job or lost a job?

Yes ○
No ○
(Skip to Item 42)

41. *(If any weeks not accounted for, ask):* What was . . . doing MOST of the remaining weeks in 1964? Was he -

All weeks accounted for in Items 34 and 38. . ○
Ill or disabled and unable to work . . . . . . ○
Taking care of home or family. . . . . ○
Going to school. . . . . . . ○
Other *(Specify)* . . . . ○

(Ask Item 42)

42. When . . . was working in 1964, did he usually work full-time or part-time?

Full-time ○ Part-time ○

43. What was . . .'s longest job in 1964? *(Compare with entry in Item 26)*

Same as Item 26 ○ *(Skip to Item 44)*
Different from Item 26. *(or Item 26 blank)* ○ *(Specify below and go to Item 44)*

A. NAME OF EMPLOYER

B. INDUSTRY

C. OCCUPATION

D. CLASS OF WORKER

P	G	O	WP
○	○	○	○

(Continue with Item 44)

FOR OFFICE USE ONLY

IND.		OCC.	
0	○ ○	0	○ ○ ○
1	○ ○	1	○ ○ ○
2	○ ○	2	○ ○ ○
3	○ ○	3	○ ○ ○
4	○ ○	4	○ ○ ○
5	○ ○	5	○ ○ ○
6	○ ○ ○	6	○ ○
7	○ ○ ○	7	○ ○
8	○ ○ ○	8	○ ○
9	○ ○ ○	9	○ ○

A	○	G	○	N	○	U	○
B	○	H	○	P	○	V	○
C	○	J	○	Q	○	W	○
D	○	K	○	R	○	X	○
E	○	L	○	S	○	Y	○
F	○	M	○	T	○	Z	○

INTERVIEWER CHECK ITEMS

44. For persons with entries in BOTH Items 34 and 38, add write-in entries of weeks, enter in box, and mark one circle. ☐

1-13 wks ○
14-26 wks ○
27-39 wks ○
40-47 wks ○
48-49 wks ○
50-52 wks ○

(Fill Item 45)

45. Is this a married, widowed, divorced, or separated woman?

Yes . . . . ○
No - Never married. . ○

(End questions)

46. Does she have any children of her own 0-13 years of age who are members of this household?

Yes - has children 0-13 ○
No children 0-13 ○

(End questions)

47. Did she work 27 weeks or more during 1964?

Yes - 27 + weeks in item 34 . . ○ *(Fill CPS-548)*
No - None or 1-26 weeks in item 34 ○

(End questions)

4. MARCH SUPPLEMENT

FILL FOR EACH HOUSEHOLD MEMBER 0 - 13 YEARS OF AGE

Budget Bureau No. 41-R 2276
Approval Expires 2-28-66

FIRST CHILD

27. LINE NO.

0	○	○
1	○	○
2	○	○
3	○	○
4		○
5		○
6		○
7		○
8		○
9		○

28. RELATIONSHIP TO HEAD OF HOUSEHOLD *(Control Card Item 15)*

(Enter relationship and mark one circle below) ☐

Child . . . . ○
Other relative. ○
Nonrelative -- own relatives in household . . ○
Nonrelative -- no own rels. in household . . ○

OFFICE USE ONLY

Family No.

1	2	3	4	5	6
○	○	○	○	○	○

Fam.Rel.: Child. . ○ Other relative ○

Type: Sl. . . . ○ Sec.F ○ Sub.F ○

29. AGE LAST BIRTHDAY *(Mark one circle only)*

0	○	7	○
1	○	8	○
2	○	9	○
3	○	10	○
4	○	11	○
5	○	12	○
6	○	13	○

31. RACE

White . . ○
Negro . . ○
Other . . ○

32. SEX

Male . . . ○
Female ○

34. Was . . . living in this house on March 1, a year ago?

No ○ *(Ask Item 35)*
Yes ○
Born after March 1, 1964 ○ *(Omit Items 35 & 36)*

If "No" in Item 34

35. Was . . . living in this same county on March 1, a year ago?

No ○ *(Ask Item 36)*
Yes ○ *(Omit Item 36)*

If "No" in Item 35

36. What State (or foreign country) was . . . living in on March 1, a year ago?

(Enter State or foreign country and mark one circle below) ☐

This State . . . . ○
Different State. ○
Abroad . . . . ○

OFFICE USE ONLY

NE	NC	S	W
○	○	○	○
C	○	N	○

SECOND CHILD

27. LINE NO.

0	○	○
1	○	○
2	○	○
3	○	○
4		○
5		○
6		○
7		○
8		○
9		○

28. RELATIONSHIP TO HEAD OF HOUSEHOLD *(Control Card Item 15)*

(Enter relationship and mark one circle below) ☐

Child . . . . ○
Other relative. ○
Nonrelative -- own relatives in household . . ○
Nonrelative -- no own rels. in household . . ○

OFFICE USE ONLY

Family No.

1	2	3	4	5	6
○	○	○	○	○	○

Fam.Rel.: Child. . ○ Other relative ○

Type: Sl. . . . ○ Sec.F ○ Sub.F ○

29. AGE LAST BIRTHDAY *(Mark one circle only)*

0	○	7	○
1	○	8	○
2	○	9	○
3	○	10	○
4	○	11	○
5	○	12	○
6	○	13	○

31. RACE

White . . ○
Negro . . ○
Other . . ○

32. SEX

Male . . . ○
Female ○

34. Was . . . living in this house on March 1, a year ago?

No ○ *(Ask Item 35)*
Yes ○
Born after March 1, 1964 ○ *(Omit Items 35 & 36)*

If "No" in Item 34

35. Was . . . living in this same county on March 1, a year ago?

No ○ *(Ask Item 36)*
Yes ○ *(Omit Item 36)*

If "No" in Item 35

36. What State (or foreign country) was . . . living in on March 1, a year ago?

(Enter State or foreign country and mark one circle below) ☐

This State . . . . ○
Different State. ○
Abroad . . . . ○

OFFICE USE ONLY

NE	NC	S	W
○	○	○	○
C	○	N	○

THIRD CHILD

27. LINE NO.

0	○	○
1	○	○
2	○	○
3	○	○
4		○
5		○
6		○
7		○
8		○
9		○

28. RELATIONSHIP TO HEAD OF HOUSEHOLD *(Control Card Item 15)*

(Enter relationship and mark one circle below) ☐

Child . . . . ○
Other relative. ○
Nonrelative -- own relatives in household . . ○
Nonrelative -- no own rels. in household . . ○

OFFICE USE ONLY

Family No.

1	2	3	4	5	6
○	○	○	○	○	○

Fam.Rel.: Child. . ○ Other relative ○

Type: Sl. . . . ○ Sec.F ○ Sub.F ○

29. AGE LAST BIRTHDAY *(Mark one circle only)*

0	○	7	○
1	○	8	○
2	○	9	○
3	○	10	○
4	○	11	○
5	○	12	○
6	○	13	○

31. RACE

White . . ○
Negro . . ○
Other . . ○

32. SEX

Male . . . ○
Female ○

34. Was . . . living in this house on March 1, a year ago?

No ○ *(Ask Item 35)*
Yes ○
Born after March 1, 1964 ○ *(Omit Items 35 & 36)*

If "No" in Item 34

35. Was . . . living in this same county on March 1, a year ago?

No ○ *(Ask Item 36)*
Yes ○ *(Omit Item 36)*

If "No" in Item 35

36. What State (or foreign country) was . . . living in on March 1, a year ago?

(Enter State or foreign country and mark one circle below) ☐

This State . . . . ○
Different State. ○
Abroad . . . . ○

OFFICE USE ONLY

NE	NC	S	W
○	○	○	○
C	○	N	○

4. MARCH SUPPLEMENT *(continued)*

FILL FOR ALL HOUSEHOLD MEMBERS 14 YEARS OLD AND OVER -- INCLUDING THOSE IN ARMED FORCES

27. LINE NUMBER

0	○	○
1	○	○
2	○	○
3	○	○
4		○
5		○
6		○
7		○
8		○
9		○

27a. POPULATION STATUS

Civilian 14+ ○

Armed Forces member ○

(AF in Control Card item 23)

28. RELATIONSHIP TO HEAD OF HOUSEHOLD
(Control Card Item 15)

(Enter relationship and mark one circle below)

Head .... ○
Wife of head ... ○
Child.......... ○
Other relative ·· ○
Nonrelative of head with own rels. in h'hld ... ○
Nonrelative of head with no own rels. in h'hld ... ○

OFFICE USE ONLY

Family No.

1	2	3	4	5	6
○	○	○	○	○	○

Fam. Rel.		Type	
Head..	○	Sec. I..	○
Wife ..	○	Sec. Fam.	○
Child..	○	Sub. Fam.	○
Other rel.	○	Pri. I..	○

29. AGE

0	○	○
1	○	○
2	○	○
3	○	○
4	○	○
5	○	○
6	○	○
7	○	○
8	○	○
9	○	○

30. MARITAL STATUS

Married--spouse present ○
Married--spouse absent in Armed Forces......... ○
Married--spouse absent--other reason ... ○
Widowed............ ○
Divorced............ ○
Separated .... ○
Never married........ ○

31. RACE

White ○
Negro ○
Other ○

32. SEX

Male.. ○ *(Fill 33)*
Female ○

33. VETERAN STATUS

Korean War .... ○
World War II.... ○
World War I.... ○
Peacetime "PA" ○
Peacetime "PB" ○
Nonveteran.... ○

34. Was . . . living in this house on March 1, a year ago?

No ○ *(Ask item 35)*
Yes ○ *(Skip to item 37)*

If "No" in 34

35. Was . . . living in this same county on March 1, a year ago?

No ○ *(Ask item 36)*
Yes ○ *(Skip to item 37)*

If "No" in 35

36. What State *(or foreign country)* **was . . . living in on March 1, a year ago?**

(Enter State or foreign country and mark circle below)

This State... ○
Different State ○
Abroad..... ○

OFFICE USE ONLY

NE	NC	S	W
○	○	○	○
C	○	N	○

EDUCATIONAL ATTAINMENT *(For all persons 14 yrs. old and over. Transcribe from CC items 24a-24b, if available, otherwise ask):*

37. What is the highest grade *(year)* **of regular school...has ever attended?**

Never attended.... ○ *(Skip to item 39)*

	1	2	3	4	5	6	7	8
Elementary school.....	○	○	○	○	○	○	○	○
High school..	○	○	○	○				
	1	2	3	4	5	6 or more		
College	○	○	○	○	○	○		

38. Did . . . finish this grade *(year)***?**

Yes ○
No ○

(If never married, skip to 40)

FOR PERSONS EVER MARRIED

39. When did . . . first get married? *(mark month, decade, and year)*

Jan.-Mar. ○
Apr.-Dec. ○

	0	1	2	3	4	5	6	7	8	9
Decade	○	○	○	○	○	○	○	○	○	○
Year	○	○	○	○	○	○	○	○	○	○

(Continue with item 40)

INCOME IN THE YEAR 1964 *(Ask for all persons 14 years of age and over except in rotations 4 and 8. Write actual amount in space provided in items 40-43 and mark appropriate circles. Mark $10,000 + circle if $10,000 or more.)*

LAST YEAR (1964) How much did . . . receive:

40. In wages or salary?

$

$10,000+	0	○	○	○	○
○	1	○	○	○	○
	2	○	○	○	○
None	3	○	○	○	○
○	4	○	○	○	○
	5	○	○	○	○
	6	○	○	○	○
	7	○	○	○	○
	8	○	○	○	○
◌	9	○	○	○	○

41. In net income from his own business or professional practice or partnership?

$

$10,000+	0	○	○	○	○
○	1	○	○	○	○
	2	○	○	○	○
None	3	○	○	○	○
○	4	○	○	○	○
	5	○	○	○	○
Lost money	6	○	○	○	○
○	7	○	○	○	○
	8	○	○	○	○
◌	9	○	○	○	○

42. In net income from his own farm?

$

$10,000+	0	○	○	○	○
○	1	○	○	○	○
	2	○	○	○	○
None	3	○	○	○	○
○	4	○	○	○	○
	5	○	○	○	○
Lost money	6	○	○	○	○
○	7	○	○	○	○
	8	○	○	○	○
◌	9	○	○	○	○

43. INCOME OTHER THAN EARNINGS. *(Ask each of items 43a, b, c, and d; add up amounts in these items and record below)*

$

$10,000+	0	○	○	○	○
○	1	○	○	○	○
	2	○	○	○	○
None	3	○	○	○	○
○	4	○	○	○	○
	5	○	○	○	○
Lost money	6	○	○	○	○
○	7	○	○	○	○
	8	○	○	○	○
◌	9	○	○	○	○

During 1964, did ... receive any of the following kinds of money income from:

43a. Social Security *(old age, survivors disability and Railroad Retirement insurance checks from the U.S. Government)*?

Yes ○ → How Much? $
No ○

43b. Dividends, interest on savings accounts or bonds, income from estates or trusts, or net rental income?

Yes ○ → How Much? $
No ○

43c. Unemployment compensation, public assistance, or welfare payments *(include aid to families with dependent children, old age assistance, and aid to the blind or totally disabled)*?

Yes ○ → How Much? $
No ○

43d. Anything else--private or government pensions, veterans' payments, annuities, regular contributions from persons not living in this household, alimony, etc.?

Yes ○ → How Much? $
No ○

5. MAY SUPPLEMENT

34. INTERVIEWER CHECK ITEM--

Mark one circle in this column and proceed accordingly.

○ Entry in Item 21 and P or G in Item 26D *(Fill 35)*

○ Entry in Item 21 and O or WP in Item 26D *(Ask 39)*

○ Entry in item 25 *(Ask 41)*

○ All other cases *(End questions for person)*

35A. INTERVIEWER CHECK: Is this person in:

Rot. 2,4,6,8 ○ *(Ask 35B)*
All other rotations .. ○ *(Ask 36)*

35B. How much does . . . usually earn per week at this job -- before deductions?

Under $40	○	$100-$149	○
$40-$59	○	$150-$199	○
$60-$79	○	$200+	○
$80-$99	○		

(Ask 36)

36. Did . . . work for more than one employer LAST WEEK?

Yes ○ *(Skip to 42)*
No ○ *(Ask 37)*

37. In addition to working for wages or salary, did . . . operate his own farm, business, or profession LAST WEEK?

Yes ○ *(Skip to 42)*
No ○ *(Ask 38)*

38. Did . . . have any other job LAST WEEK at which he did not work at all?
(If "Yes," indicate whether paid for time off.)

Yes-Paid ... ○ } *(Skip to 42)*
Yes-Not paid ○ }
No......... ○ *(Skip to 45)*

39. In addition to this work, did . . . do any work for wages or salary LAST WEEK?

Yes ○ *(Skip to 42)*
No ○ *(Ask 40)*

40. Did . . . have a job LAST WEEK at which he did not work at all?
(If "Yes," indicate whether paid for time off.)

Yes-Paid ... ○ } *(Skip to 42)*
Yes-Not paid ○ }
No......... ○ *(End questions for person)*

41. In addition to this job, did . . . have some other job LAST WEEK at which he did not work at all?

Yes ○ *(Ask 42)*
No ○ *(End questions for person)*

DESCRIBE 2ND JOB IF "YES" IN ANY OF ITEMS 36-41

NOTE Mark whether second job is same as or different from job in item 26B-D. Describe if different.

42. What was . . .'s second job LAST WEEK?

Same as 26B-D .... ○ *(Skip to 43)*
Different from 26B-D ○ *(Describe below and go to 43)*

A. NAME OF EMPLOYER

B. INDUSTRY

C. OCCUPATION

D. CLASS OF WORKER

P ○
G ○
O ○
WP ○

43. How many hours did . . . work at his second job LAST WEEK?

0	○	○
1	○	○
2	○	○
3	○	○
4	○	○
5	○	○
6	○	○
7	○	○
8	○	○
9	○	○

(Ask 44)

44. How many hours did . . . work at his principal job (Item 26) LAST WEEK?

0	○	○
1	○	○
2	○	○
3	○	○
4	○	○
5	○	○
6	○	○
7	○	○
8	○	○
9	○	○

(End questions for person)

INDUSTRY

0	○	○		A	○	G	○
1	○	○		B	○	H	○
2	○	○		C	○	J	○
3	○	○		D	○	K	○
4	○	○		E	○	L	○
5	○	○		F	○	M	○
6	○	○	○				
7	○	○	○				
8	○	○	○				
9	○	○	○				

OCCUPATION

0	○	○	○	N	○	U	○
1	○	○	○	P	○	V	○
2	○	○	○	Q	○	W	○
3	○	○	○	R	○	X	○
4	○	○	○	S	○	Y	○
5	○	○	○	T	○	Z	○
6	○	○					
7	○	○					
8	○	○					
9	○	○					

45. INT. CHECK ITEM--Fill if "No" in item 38
(Transcribe from item 21)

40 hours or less *(Including NA's)* ○ *(End questions for person)*
41 hours or more ○ *(Ask 46 and 47)*

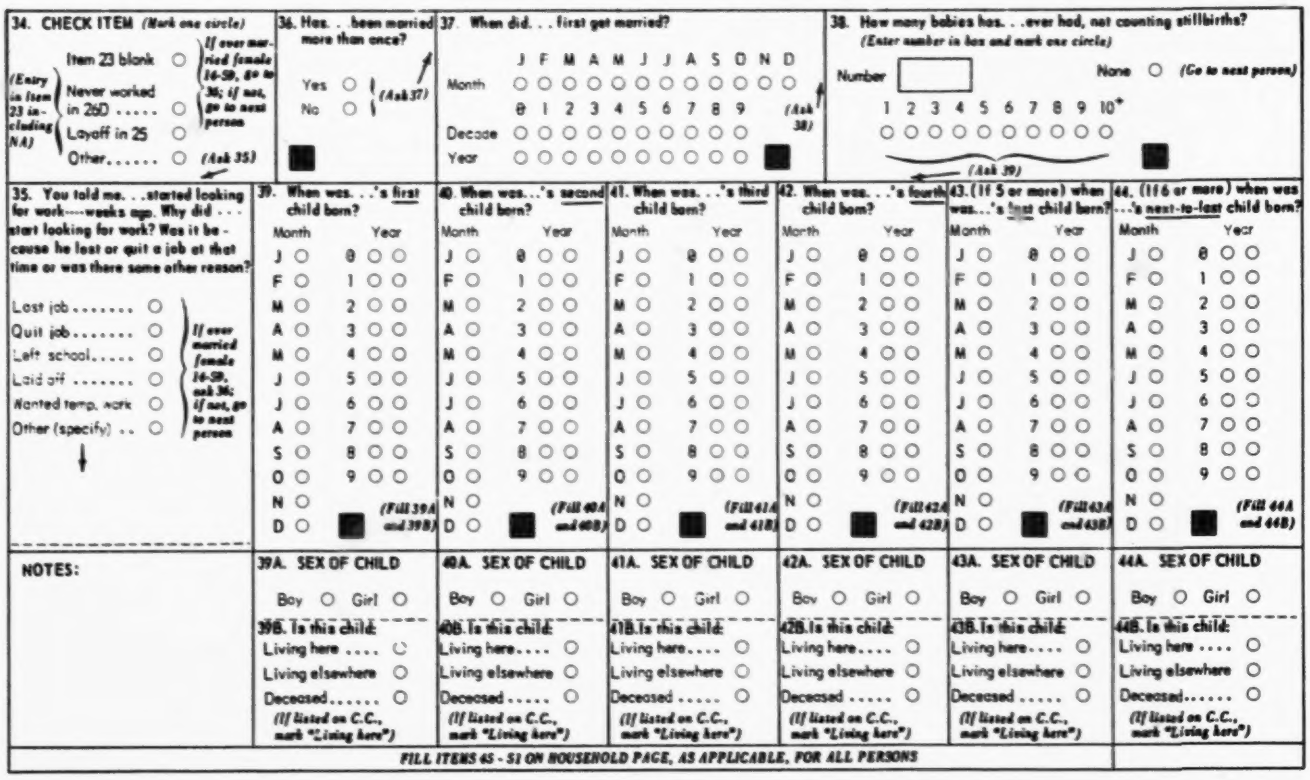
46. Did . . . get a higher rate of pay, like time and a half or double time, for the hours he worked over 40?

Yes ○ } *(Ask 47)*
No ○ }

47. Does . . . usually work more than 40 hours a week?

Yes ○ } *(End questions for person)*
No ○ }

6. JUNE SUPPLEMENT



34. CHECK ITEM *(Mark one circle)*

(Entry in Item 23 including NA)
- Item 23 blank ○
- Never worked in 26D ..... ○
- Layoff in 25 ○
- Other...... ○ *(Ask 35)*

If ever married female 14-59, go to 36; if not, go to next person

36. Has. . .been married more than once?

Yes ○
No ○
(Ask 37)

37. When did. . .first get married?

	J	F	M	A	M	J	J	A	S	O	N	D
Month	○	○	○	○	○	○	○	○	○	○	○	○

	0	1	2	3	4	5	6	7	8	9
Decade	○	○	○	○	○	○	○	○	○	○
Year	○	○	○	○	○	○	○	○	○	○

(Ask 38)

38. How many babies has. . .ever had, not counting stillbirths?
(Enter number in box and mark one circle)

Number ☐

None ○ *(Go to next person)*

1	2	3	4	5	6	7	8	9	10+
○	○	○	○	○	○	○	○	○	○

(Ask 39)

35. You told me. . .started looking for work----weeks ago. Why did . . . start looking for work? Was it because he lost or quit a job at that time or was there some other reason?

- Lost job .... ○
- Quit job .... ○
- Left school ..... ○
- Laid off .... ○
- Wanted temp. work ○
- Other (specify) .. ○

If ever married female 14-59, ask 36; if not, go to next person

39. When was. . .'s first child born?

40. When was. . .'s second child born?

41. When was. . .'s third child born?

42. When was. . .'s fourth child born?

43. (If 5 or more) when was. . .'s last child born?

44. (If 6 or more) when was. . .'s next-to-last child born?

Each of items 39–44:

Month		Year		
J	○	0	○	○
F	○	1	○	○
M	○	2	○	○
A	○	3	○	○
M	○	4	○	○
J	○	5	○	○
J	○	6	○	○
A	○	7	○	○
S	○	8	○	○
O	○	9	○	○
N	○			
D	○			

(Fill 39A and 39B) *(Fill 40A and 40B)* *(Fill 41A and 41B)* *(Fill 42A and 42B)* *(Fill 43A and 43B)* *(Fill 44A and 44B)*

NOTES:

39A. SEX OF CHILD / **40A. SEX OF CHILD** / **41A. SEX OF CHILD** / **42A. SEX OF CHILD** / **43A. SEX OF CHILD** / **44A. SEX OF CHILD**

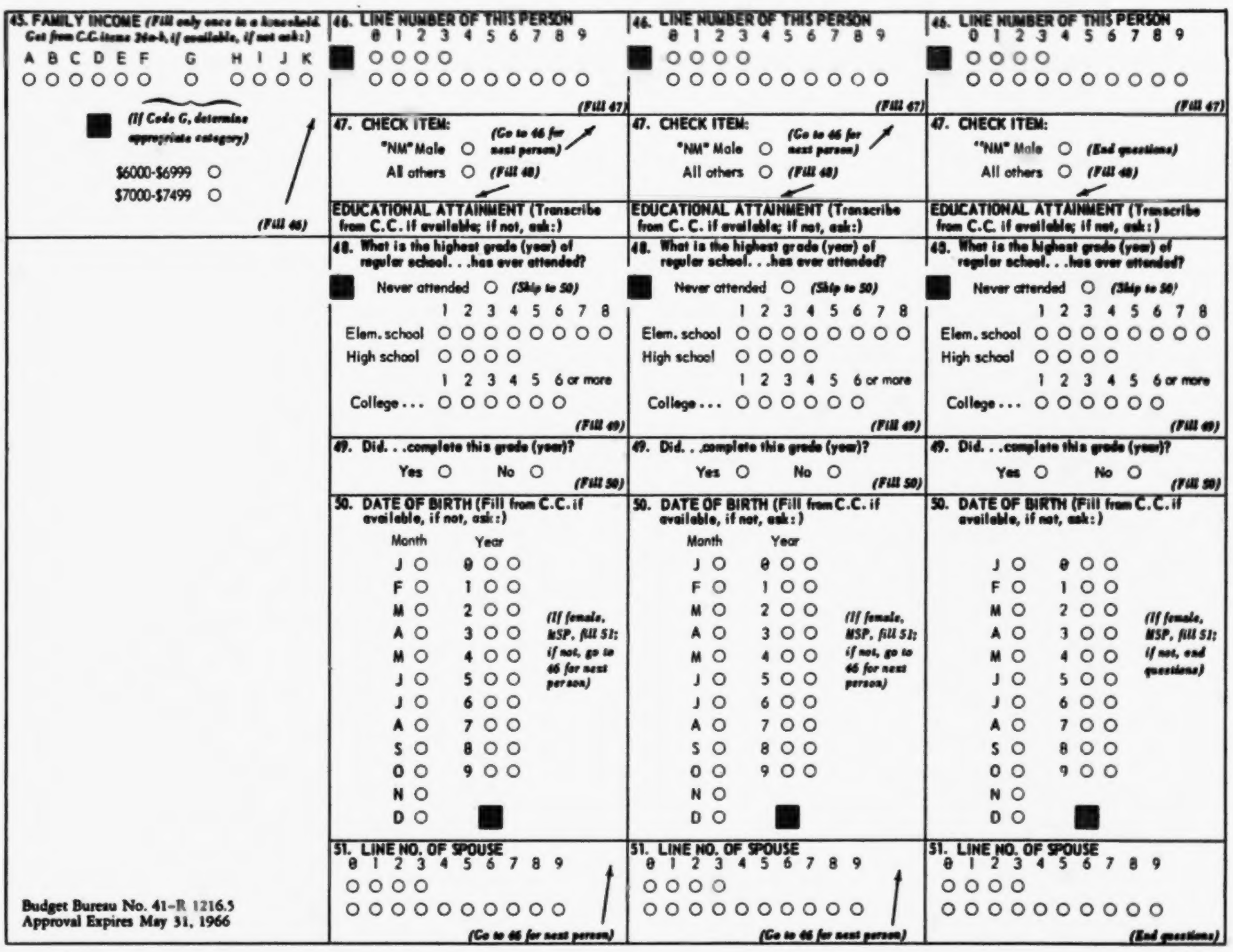
Boy ○ Girl ○

39B. Is this child: / **40B. Is this child:** / **41B. Is this child:** / **42B. Is this child:** / **43B. Is this child:** / **44B. Is this child:**

- Living here .... ○
- Living elsewhere ○
- Deceased...... ○

(If listed on C.C., mark "Living here")

FILL ITEMS 45 - 51 ON HOUSEHOLD PAGE, AS APPLICABLE, FOR ALL PERSONS



45. FAMILY INCOME *(Fill only once in a household. Get from C.C. items 26a-b, if available, if not ask:)*

A	B	C	D	E	F	G	H	I	J	K
○	○	○	○	○	○	○	○	○	○	○

(If Code G, determine appropriate category)

$6000-$6999 ○
$7000-$7499 ○

(Fill 46)

Budget Bureau No. 41-R 1216.5
Approval Expires May 31, 1966

(Each of the three person columns:)

46. LINE NUMBER OF THIS PERSON

0	1	2	3	4	5	6	7	8	9
○	○	○	○						
○	○	○	○	○	○	○	○	○	○

(Fill 47)

47. CHECK ITEM:

"NM" Male ○ *(Go to 46 for next person)* — third column: *(End questions)*
All others ○ *(Fill 48)*

EDUCATIONAL ATTAINMENT (Transcribe from C.C. if available; if not, ask:)

48. What is the highest grade (year) of regular school. . .has ever attended?

Never attended ○ *(Skip to 50)*

	1	2	3	4	5	6	7	8
Elem. school	○	○	○	○	○	○	○	○
High school	○	○	○	○				

	1	2	3	4	5	6 or more
College . . .	○	○	○	○	○	○

(Fill 49)

49. Did. . .complete this grade (year)?

Yes ○ No ○ *(Fill 50)*

50. DATE OF BIRTH (Fill from C.C. if available, if not, ask:)

Month		Year		
J	○	0	○	○
F	○	1	○	○
M	○	2	○	○
A	○	3	○	○
M	○	4	○	○
J	○	5	○	○
J	○	6	○	○
A	○	7	○	○
S	○	8	○	○
O	○	9	○	○
N	○			
D	○			

(If female, MSP, fill 51; if not, go to 46 for next person) — third column: *(If female, MSP, fill 51; if not, end questions)*

51. LINE NO. OF SPOUSE

0	1	2	3	4	5	6	7	8	9
○	○	○	○						
○	○	○	○	○	○	○	○	○	○

(Go to 46 for next person) — third column: *(End questions)*

7. OCTOBER SUPPLEMENT

FILL FOR EACH HOUSEHOLD MEMBER 3-13 YEARS OF AGE

FIRST CHILD

27. LINE NUMBER

Tens 0 1 2 3
Units 0 1 2 3 4 5 6 7 8 9

29. AGE LAST BIRTHDAY

3 4 5 6 8 7 9 10 11 12 13
○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○

31. RACE
White ○
Negro ○
Other ○

32. SEX
Male ○
Female ○

34. Is . . . attending or enrolled in school? Yes ○ No ○ *(End of questions)*
(Including nursery school and kindergarten.)

35. What level of school?
Nursery, full day . . . . ○
Nursery, part day . . . . ○ *(skip to 37)*
Kindergarten . . . . ○
Elementary . . . . ○
High school . . . . ○ *(Ask 36 and 37)*
Special school . . . . . . ○ *(Specify in notes)*

36. What grade (or year) is . . . attending?
(e.g., 7 for 7th grade elem.; 3 for 3d year H.S.)
1 2 3 4 5 6 7 8
○ ○ ○ ○ ○ ○ ○ ○

37. Is it a public or a private school?
Public ○ Private ○ *(Include parochial)*

38. *(If elem. 6 - H.S. 2 in 35 and 36)*
Is it a Junior High School?
Yes ○ No ○

SECOND CHILD

27. LINE NUMBER

Tens 0 1 2 3
Units 0 1 2 3 4 5 6 7 8 9

29. AGE LAST BIRTHDAY

3 4 5 6 7 8 9 10 11 12 13
○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○

31. RACE
White ○
Negro ○
Other ○

32. SEX
Male ○
Female ○

34. Is . . . attending or enrolled in school? Yes ○ No ○ *(End of questions)*
(Including nursery school and kindergarten.)

35. What level of school?
Nursery, full day . . . . ○
Nursery, part day . . . . ○ *(Skip to 37)*
Kindergarten . . . . ○
Elementary . . . . ○
High school . . . . ○ *(Ask 36 and 37)*
Special school . . . . . . ○ *(Specify in notes)*

36. What grade (or year) is . . . attending?
(e.g., 7 for 7th grade elem.; 3 for 3d year H.S.)
1 2 3 4 5 6 7 8
○ ○ ○ ○ ○ ○ ○ ○

37. Is it a public or a private school?
Public ○ Private ○ *(Include parochial)*

38. *(If elem. 6 - H.S. 2 in 35 and 36)*
Is it a Junior High School?
Yes ○ No ○

THIRD CHILD

27. LINE NUMBER

Tens 0 1 2 3
Units 0 1 2 3 4 5 6 7 8 9

29. AGE LAST BIRTHDAY

3 4 5 6 7 8 9 10 11 12 13
○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○

31. RACE
White ○
Negro ○
Other ○

32. SEX
Male ○
Female ○

34. Is . . . attending or enrolled in school? Yes ○ No ○ *(End of questions)*
(Including nursery school and kindergarten.)

35. What level of school?
Nursery, full day . . . . ○
Nursery, part day . . . . ○ *(Skip to 37)*
Kindergarten . . . . ○
Elementary . . . . ○
High school . . . . ○ *(Ask 36 and 37)*
Special school . . . . . . ○ *(Specify in notes)*

36. What grade (or year) is . . . attending?
(e.g., 7 for 7th grade elem.; 3 for 3d year H.S.)
1 2 3 4 5 6 7 8
○ ○ ○ ○ ○ ○ ○ ○

37. Is it a public or a private school?
Public ○ Private ○ *(Include parochial)*

38. *(If elem. 6 - H.S. 2 in 35 and 36)*
Is it a Junior High School?
Yes ○ No ○

FOURTH CHILD

27. LINE NUMBER

Tens 0 1 2 3
Units 0 1 2 3 4 5 6 7 8 9

29. AGE LAST BIRTHDAY

3 4 5 6 7 8 9 10 11 12 13
○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○

31. RACE
White ○
Negro ○
Other ○

32. SEX
Male ○
Female ○

34. Is . . . attending or enrolled in school? Yes ○ No ○ *(End of questions)*
(Including nursery school and kindergarten.)

35. What level of school?
Nursery, full day . . . . ○
Nursery, part day . . . . ○ *(Skip to 37)*
Kindergarten . . . . ○
Elementary . . . . ○
High school . . . . ○ *(Ask 36 and 37)*
Special school . . . . . . ○ *(Specify in notes)*

36. What grade (or year) is . . . attending?
(e.g., 7 for 7th grade elem.; 3 for 3d year H.S.)
1 2 3 4 5 6 7 8
○ ○ ○ ○ ○ ○ ○ ○

37. Is it a public or a private school?
Public ○ Private ○ *(Include parochial)*

38. *(If elem. 6 - H.S. 2 in 35 and 36)*
Is it a Junior High School?
Yes ○ No ○

46. TOTAL NUMBER OF HOUSEHOLD MEMBERS UNDER 14 YEARS OF AGE.
(Fill from Control Card)

BOYS | GIRLS

47. TENURE
(Transcribe from Control Card Item 9.)
Owned or being bought ○
Rented ○
No cash rent ○

ALL ROTATIONS

48. Total family income
(Transcribe from CP-21 for rotations 3, 4, 7, 8; for all other rotations, transcribe from Control Card item 33.)

A ○	G ○
B ○	H ○
C ○	I ○
D ○	J ○
E ○	K ○
F ○	

---

34. Is . . . attending or enrolled in school?
Yes
○ *(Ask 35*
No
○ *(Skip to 39)*
35+
○ *(End of questions)*

If "Yes" in Item 34)

35. **What level of school?**
(If "College," ask if full-time or part-time)
Elementary . . . . ○
High School . . . . ○
College, full-time . . . ○ *(Ask 36 and 37)*
College, part-time . . . ○
Special School . . . . . . ○
(Specify below and skip to 39)

36. **What grade (or year) is . . . attending?**
(e.g., 7 for 7th grade elem.; 3 for 3d year H.S.)
1 ○
2 ○
3 ○
4 ○
5 ○
6 ○
7 ○
8 ○

37. **Is it a public or a private school?**
Public. ○
Private . . ○ *(Include parochial)*
(Ask 38, if applicable, then go to 39.)

38. *(If elem. 6 - H.S. 2 in 35 and 36.)*
Is it a Junior High School?
Yes ○
No ○
(Go to 39)

39. INTERVIEWER CHECK ITEM
(Fill for all persons 14 - 34 years old)
(Mark appropriate circle below for this person)
Enrolled - College, full- or part-time in 35 . . . . ○ *(Ask 40)*
Not enrolled in school ("No" in 34) and 14 - 24 years old . . . ○ *(Ask 41)*
Enrolled in Special School in 35 and 14 - 24 years old . . . . ○ *(Ask 41)*
All other persons . . . . ○ *(End of questions)*

If "College, full-time" or "College, part-time" in Item 35)

40. **When did . . . graduate from High School?**
June, 1964 . . . . ○
January, 1964 . . . . . ○
1963 . . . . ○
1962 . . . . ○
1961 . . . . ○
1960 . . . . ○
1959 or earlier . . . . ○
(End of questions)

If 14 - 24 years old with "No" in 34) or "Special School" in 35

41. **Is . . . a High School graduate?**
Yes ○ No ○ *(Ask 43)*

42. **When did . . . graduate from High School?**
June, 1964 . . ○
January, 1964 ○
1963 . . . . . . ○
1962 . . . . . . ○
1961 . . . . . . ○
1960 . . . . ○
1959 or earlier ○ *(End of questions)*

If "No" or entry unknown in Item 41

43. **Was . . . attending or enrolled in regular school in October 1963, that is, in October of last year?**
Yes ○ No ○ *(Skip to 45)*

44. **What grade (year) of regular school was . . . attending or enrolled in when . . . left school?**
High School . . . . . ○
(grades 9-12)
Elementary School . . . . ○
(grades 1-8)
(Ask Item 45)

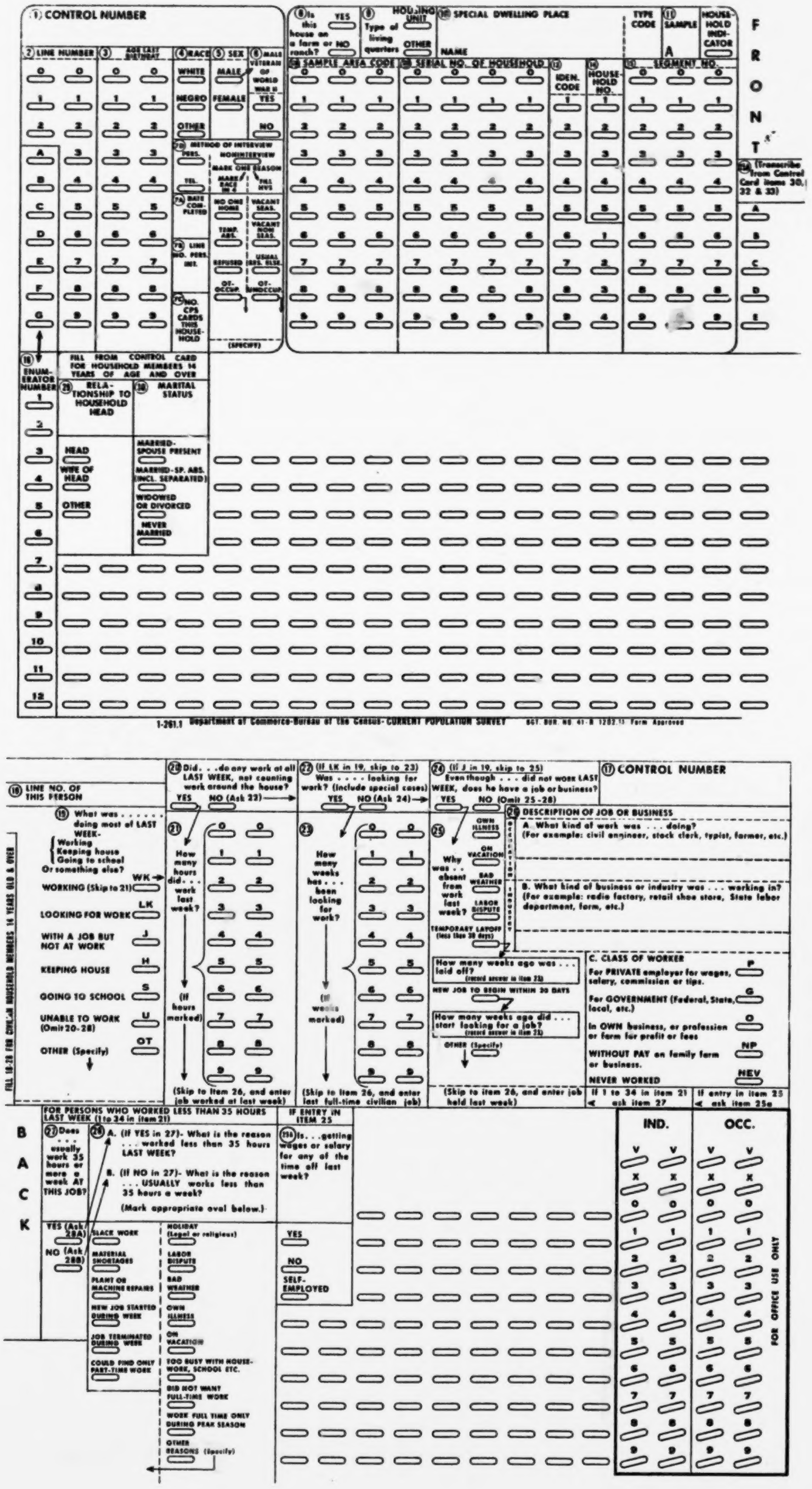
45. **What year did . . . last attend regular school?**
1964 . . . . ○
1963 . . . . ○
1962 . . . . ○
1961 . . . . ○
1960 . . . . ○
1959 or earlier . . . ○
Never attended . . . ○

NOTES

Budget Bureau No. 41-R 1221.18
Approval Expires 9 - 30 - 65

8a. CPS SCHEDULE

CPS Schedule in use from Feb. 1952 to Oct. 1961
(for household head)



FRONT

① CONTROL NUMBER

② LINE NUMBER
○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9
○ A ○ B ○ C ○ D ○ E ○ F ○ G

③ AGE LAST BIRTHDAY
○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9

④ RACE
○ WHITE
○ NEGRO
○ OTHER

⑤ SEX
○ MALE
○ FEMALE

⑥ MALE VETERAN OF WORLD WAR II
○ YES
○ NO

⑦ METHOD OF INTERVIEW
○ PERS.
○ TEL.
NONINTERVIEW ○
MARK ONE REASON
MARK RACE IN 4
FILL HVS

⑦A DATE COMPLETED

⑦B LINE NO. PERS. INT.

⑦C NO. CPS CARDS THIS HOUSEHOLD

○ NO ONE HOME
○ TEMP. ABS.
○ REFUSED
○ OT.-OCCUP.
○ VACANT SEAS.
○ VACANT NON SEAS.
○ USUAL RES. ELSE.
○ OT.-UNOCCUP.
(SPECIFY)

⑧ Is this house on a farm or ranch?
○ YES
○ NO

⑨ Type of living quarters
○ HOUSING UNIT
○ OTHER

⑩ SPECIAL DWELLING PLACE
NAME
TYPE CODE

⑪ SAMPLE
A

HOUSEHOLD INDICATOR ○

⑫ SAMPLE AREA CODE
○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9

⑬ SERIAL NO. OF HOUSEHOLD
○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9

⑫ IDEN. CODE
○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9

⑭ HOUSEHOLD NO.
○ 1 ○ 2 ○ 3 ○ 4 ○ 5

○ 1 ○ 2 ○ 3 ○ 4

⑮ SEGMENT NO.
○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9

⑯A (Transcribe from Control Card items 30, 32 & 33)
○ A ○ B ○ C ○ D ○ E

⑯ ENUMERATOR NUMBER
○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10 ○ 11 ○ 12

FILL FROM CONTROL CARD FOR HOUSEHOLD MEMBERS 14 YEARS OF AGE AND OVER

㉙ RELATIONSHIP TO HOUSEHOLD HEAD
○ HEAD
○ WIFE OF HEAD
○ OTHER

㉚ MARITAL STATUS
○ MARRIED-SPOUSE PRESENT
○ MARRIED-SP. ABS. (INCL. SEPARATED)
○ WIDOWED OR DIVORCED
○ NEVER MARRIED

1-261.1 Department of Commerce-Bureau of the Census-CURRENT POPULATION SURVEY BGT. BUR. NO 41-R 1202.13 Form Approved

BACK

⑱ LINE NO. OF THIS PERSON

FILL 19-28 FOR CIVILIAN HOUSEHOLD MEMBERS 14 YEARS OLD & OVER

⑲ What was . . . . . . doing most of LAST WEEK-
Working
Keeping house
Going to school
Or something else?
WORKING (Skip to 21) ○ WK
LOOKING FOR WORK ○ LK
WITH A JOB BUT NOT AT WORK ○ J
KEEPING HOUSE ○ H
GOING TO SCHOOL ○ S
UNABLE TO WORK (Omit 20-28) ○ U
OTHER (Specify) ○ OT

⑳ Did. . .do any work at all LAST WEEK, not counting work around the house?
○ YES
○ NO (Ask 22)

㉑ How many hours did. . . work last week?
○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9
(If hours marked)
(Skip to Item 26, and enter job worked at last week)

㉒ (If LK in 19, skip to 23)
Was . . . . looking for work? (Include special cases)
○ YES
○ NO (Ask 24)

㉓ How many weeks has . . . been looking for work?
○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9
(If weeks marked)
(Skip to Item 26, and enter last full-time civilian job)

㉔ (If J in 19, skip to 25)
Even though . . . did not work LAST WEEK, does he have a job or business?
○ YES
○ NO (Omit 25-28)

㉕ Why was . . absent from work last week?
○ OWN ILLNESS
○ ON VACATION
○ BAD WEATHER
○ LABOR DISPUTE
○ TEMPORARY LAYOFF (less than 30 days)
How many weeks ago was . . . laid off? (record answer in item 23)
○ NEW JOB TO BEGIN WITHIN 30 DAYS
How many weeks ago did . . . start looking for a job? (record answer in item 23)
○ OTHER (Specify)
(Skip to item 26, and enter job held last week)

⑰ CONTROL NUMBER

㉖ DESCRIPTION OF JOB OR BUSINESS
OCCUPATION
A. What kind of work was . . . doing?
(For example: civil engineer, stock clerk, typist, farmer, etc.)
INDUSTRY
B. What kind of business or industry was . . . working in?
(For example: radio factory, retail shoe store, State labor department, farm, etc.)
C. CLASS OF WORKER
For PRIVATE employer for wages, salary, commission or tips. ○ P
For GOVERNMENT (Federal, State, local, etc.) ○ G
In OWN business, or profession or farm for profit or fees ○ O
WITHOUT PAY on family farm or business. ○ NP
NEVER WORKED ○ NEV

If 1 to 34 in item 21 ask item 27

If entry in item 25 ask item 25a

FOR PERSONS WHO WORKED LESS THAN 35 HOURS LAST WEEK (1 to 34 in item 21)

㉗ Does . . . usually work 35 hours or more a week AT THIS JOB?
○ YES (Ask 28A)
○ NO (Ask 28B)

㉘ A. (If YES in 27)- What is the reason . . . worked less than 35 hours LAST WEEK?
B. (If NO in 27)- What is the reason . . . USUALLY works less than 35 hours a week?
(Mark appropriate oval below.)
○ SLACK WORK
○ MATERIAL SHORTAGES
○ PLANT OR MACHINE REPAIRS
○ NEW JOB STARTED DURING WEEK
○ JOB TERMINATED DURING WEEK
○ COULD FIND ONLY PART-TIME WORK
○ HOLIDAY (Legal or religious)
○ LABOR DISPUTE
○ BAD WEATHER
○ OWN ILLNESS
○ ON VACATION
○ TOO BUSY WITH HOUSEWORK, SCHOOL ETC.
○ DID NOT WANT FULL-TIME WORK
○ WORK FULL TIME ONLY DURING PEAK SEASON
○ OTHER REASONS (Specify)

IF ENTRY IN ITEM 25
㉕a Is. . .getting wages or salary for any of the time off last week?
○ YES
○ NO
○ SELF-EMPLOYED

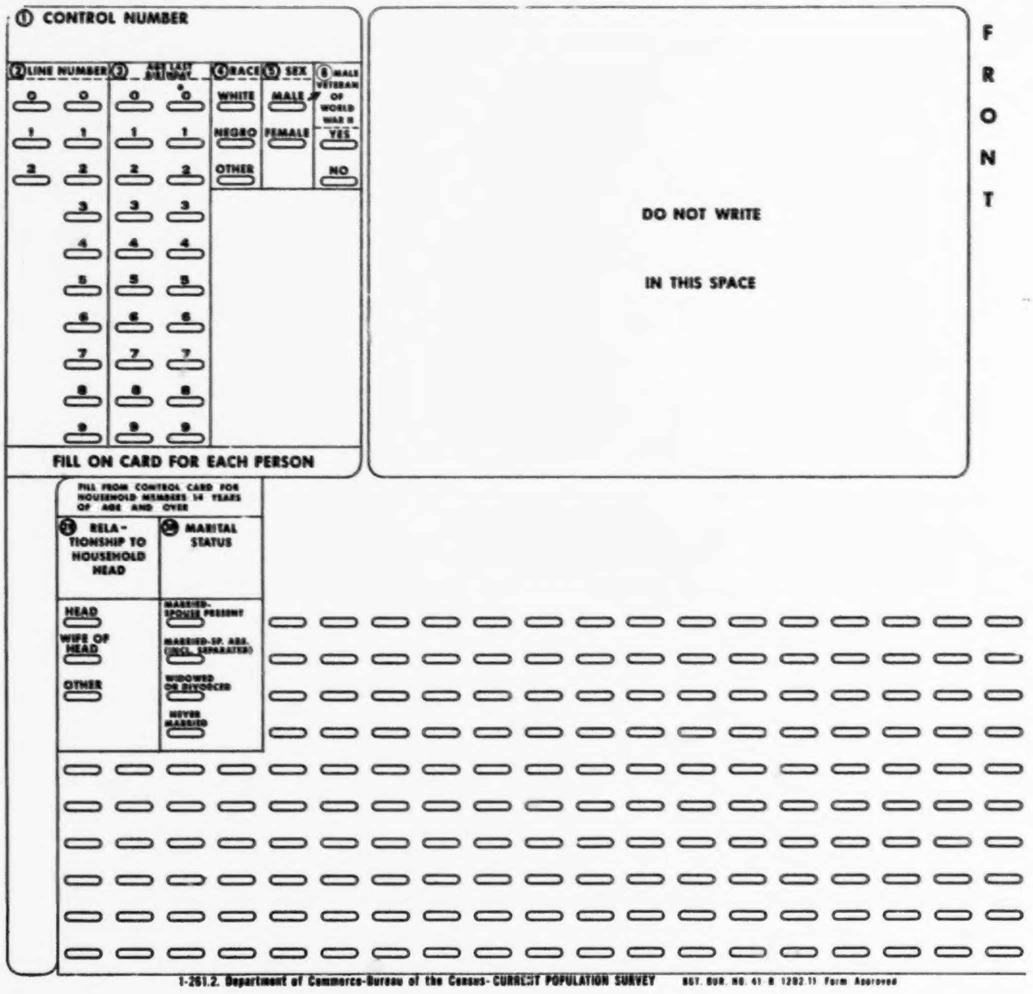
IND.
○ V ○ X ○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9

OCC.
○ V ○ X ○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9

FOR OFFICE USE ONLY

8b. CPS SCHEDULE

CPS Schedule in use from Feb. 1952 to Oct. 1961
(for other household member than household head, 14 yrs. and over)



F R O N T

① CONTROL NUMBER

② LINE NUMBER: 0 0 / 1 1 / 2 2 / 3 / 4 / 5 / 6 / 7 / 8 / 9

③ [illegible]: 0 0 / 1 1 / 2 2 / 3 3 / 4 4 / 5 5 / 6 6 / 7 7 / 8 8 / 9 9

④ RACE: WHITE ○ NEGRO ○ OTHER ○

⑤ SEX: MALE ○ FEMALE ○

⑥ MALE VETERAN OF WORLD WAR II: YES ○ NO ○

FILL ON CARD FOR EACH PERSON

DO NOT WRITE

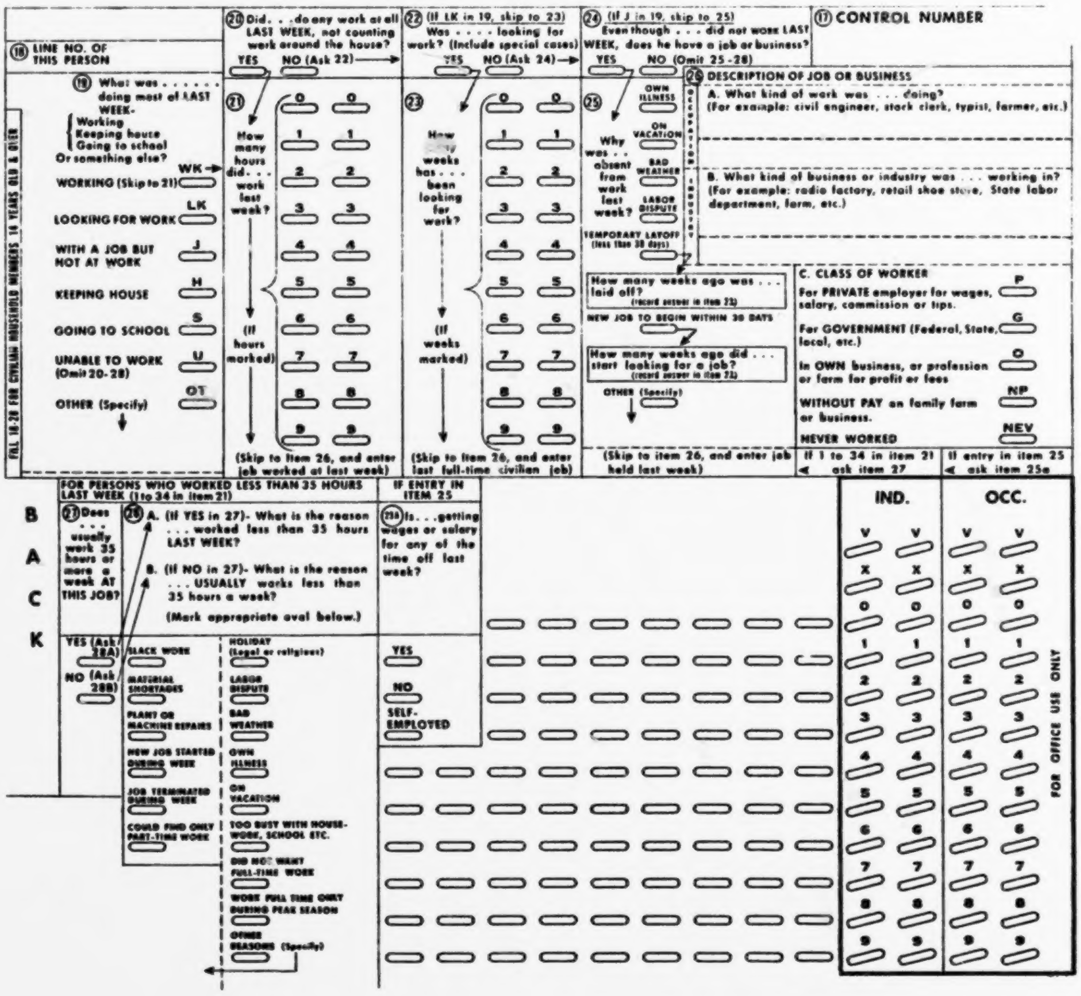
IN THIS SPACE

FILL FROM CONTROL CARD FOR HOUSEHOLD MEMBERS 14 YEARS OF AGE AND OVER

RELATIONSHIP TO HOUSEHOLD HEAD: HEAD ○ WIFE OF HEAD ○ OTHER ○

MARITAL STATUS: MARRIED-SPOUSE PRESENT ○ MARRIED-SP. ABS. (INCL. SEPARATED) ○ WIDOWED OR DIVORCED ○ NEVER MARRIED ○

1-261.2. Department of Commerce-Bureau of the Census- CURRENT POPULATION SURVEY BGT. BUR. NO. 41-R 1202.11 Form Approved



B A C K

⑱ LINE NO. OF THIS PERSON

⑲ What was . . . . . . doing most of LAST WEEK- Working, Keeping house, Going to school, Or something else?

- WORKING (Skip to 21) WK ○
- LOOKING FOR WORK LK ○
- WITH A JOB BUT NOT AT WORK J ○
- KEEPING HOUSE H ○
- GOING TO SCHOOL S ○
- UNABLE TO WORK (Omit 20-28) U ○
- OTHER (Specify) OT ○

FILL 18-28 FOR CIVILIAN HOUSEHOLD MEMBERS 14 YEARS OLD & OVER

⑳ Did. . . do any work at all LAST WEEK, not counting work around the house? YES ○ NO (Ask 22) ○

㉑ How many hours did . . . work last week? 0–9 0–9 (If hours marked) (Skip to Item 26, and enter job worked at last week)

㉒ (If LK in 19, skip to 23) Was . . . . looking for work? (Include special cases) YES ○ NO (Ask 24) ○

㉓ How many weeks has . . . been looking for work? 0–9 0–9 (If weeks marked) (Skip to Item 26, and enter last full-time civilian job)

㉔ (If J in 19, skip to 25) Even though . . . did not work LAST WEEK, does he have a job or business? YES ○ NO (Omit 25-28) ○

㉕ Why was . . . absent from work last week? OWN ILLNESS ○ ON VACATION ○ BAD WEATHER ○ LABOR DISPUTE ○ TEMPORARY LAYOFF (less than 30 days) ○

How many weeks ago was . . . laid off? (record answer in item 23)

NEW JOB TO BEGIN WITHIN 30 DAYS ○

How many weeks ago did . . . start looking for a job? (record answer in item 23)

OTHER (Specify) ○

(Skip to item 26, and enter job held last week)

⑰ CONTROL NUMBER

㉖ DESCRIPTION OF JOB OR BUSINESS

OCCUPATION — A. What kind of work was . . . doing? (For example: civil engineer, stock clerk, typist, farmer, etc.)

INDUSTRY — B. What kind of business or industry was . . . working in? (For example: radio factory, retail shoe store, State labor department, farm, etc.)

C. CLASS OF WORKER

- For PRIVATE employer for wages, salary, commission or tips. P ○
- For GOVERNMENT (Federal, State, local, etc.) G ○
- In OWN business, or profession or farm for profit or fees O ○
- WITHOUT PAY on family farm or business. NP ○
- NEVER WORKED NEV ○

If 1 to 34 in item 21 ◄ ask item 27

If entry in item 25 ◄ ask item 25a

FOR PERSONS WHO WORKED LESS THAN 35 HOURS LAST WEEK (1 to 34 in item 21)

㉗ Does . . . usually work 35 hours or more a week AT THIS JOB? YES (Ask 28A) ○ NO (Ask 28B) ○

㉘ A. (If YES in 27)- What is the reason . . . worked less than 35 hours LAST WEEK?

B. (If NO in 27)- What is the reason . . . USUALLY works less than 35 hours a week?

(Mark appropriate oval below.)

- SLACK WORK ○
- MATERIAL SHORTAGES ○
- PLANT OR MACHINE REPAIRS ○
- NEW JOB STARTED DURING WEEK ○
- JOB TERMINATED DURING WEEK ○
- COULD FIND ONLY PART-TIME WORK ○
- HOLIDAY (Legal or religious) ○
- LABOR DISPUTE ○
- BAD WEATHER ○
- OWN ILLNESS ○
- ON VACATION ○
- TOO BUSY WITH HOUSEWORK, SCHOOL ETC. ○
- DID NOT WANT FULL-TIME WORK ○
- WORK FULL TIME ONLY DURING PEAK SEASON ○
- OTHER REASONS (Specify) ○

IF ENTRY IN ITEM 25

㉕ᴬ Is . . . getting wages or salary for any of the time off last week? YES ○ NO ○ SELF-EMPLOYED ○

IND. OCC. V V V V / X X X X / 0 0 0 0 / 1 1 1 1 / 2 2 2 2 / 3 3 3 3 / 4 4 4 4 / 5 5 5 5 / 6 6 6 6 / 7 7 7 7 / 8 8 8 8 / 9 9 9 9

FOR OFFICE USE ONLY